

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 18:01:41 ; Search time 97.89 Seconds
(without alignments)
7107.370 Million cell updates/sec

Title: US-09-697-089-3
Perfect score: 3072
Sequence: 1 atgaattcataagacaa.....ctttaactagtaactgct 3072

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 11328999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued_patents_NA :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	0.6	1094	2 US-08-870-518-34	Sequence 34, App1
2	18	0.6	1515	3 US-08-747-221B-16	Sequence 16, App1
3	18	0.6	1515	3 US-08-747-221B-17	Sequence 17, App1
4	18	0.6	1515	4 US-09-005-051-16	Sequence 16, App1
5	18	0.6	1515	4 US-09-005-051-17	Sequence 17, App1
6	18	0.6	1611	6 5213972-6	Patent No. 5213972
7	18	0.6	1650	3 US-08-747-221B-60	Sequence 60, App1
8	18	0.6	1650	3 US-08-747-221B-61	Sequence 61, App1
9	18	0.6	1650	4 US-09-005-051-60	Sequence 60, App1
10	18	0.6	1650	4 US-09-005-051-61	Sequence 61, App1
11	18	0.6	1721	3 US-08-857-213-2	Sequence 2, App1
12	18	0.6	1800	1 US-08-139-937-11	Sequence 11, App1
13	18	0.6	1800	1 PCT-US93-11310-11	Sequence 11, App1
14	18	0.6	1982	3 US-08-747-221B-13	Sequence 13, App1
15	18	0.6	1982	3 US-08-747-221B-15	Sequence 15, App1
16	18	0.6	1982	4 US-09-005-051-13	Sequence 13, App1
17	18	0.6	1982	4 US-09-005-051-15	Sequence 15, App1
18	18	0.6	2144	3 US-08-747-221B-57	Sequence 57, App1
19	18	0.6	2144	3 US-08-747-221B-59	Sequence 59, App1
20	18	0.6	2144	4 US-09-005-051-57	Sequence 57, App1
21	18	0.6	2144	4 US-09-005-051-59	Sequence 59, App1
22	18	0.6	3000	4 US-09-192-104-1	Sequence 1, App1
23	18	0.6	3000	4 US-09-543-446-1	Sequence 1, App1
24	18	0.6	4982	3 US-08-699-103B-1	Sequence 1, App1
25	18	0.6	290	4 US-09-037-990B-27	Sequence 27, App1
26	17	0.6	383	4 US-08-991-789A-171	Sequence 171, App1
27	17	0.6	661	4 US-09-037-990B-8	Sequence 8, App1

28	17	0.6	685	4 US-08-998-416-818	Sequence 818, App
29	17	0.6	1536	2 US-09-318-443-5	Sequence 5, App1
30	17	0.6	1563	2 US-08-743-637B-163	Sequence 163, App
31	17	0.6	1682	4 US-09-318-443-7	Sequence 7, App1
32	17	0.6	1965	4 US-09-178-252-26	Sequence 26, App1
33	17	0.6	2035	4 US-08-960-780-10	Sequence 10, App1
34	17	0.6	2035	4 US-09-073-898-10	Sequence 10, App1
35	17	0.6	2521	4 US-09-115-446-1	Sequence 1, App1
36	17	0.6	2526	4 US-09-115-446-5	Sequence 5, App1
37	17	0.6	4165	1 US-08-095-737-1	Sequence 1, App1
38	17	0.6	4165	1 US-08-480-145-1	Sequence 1, App1
39	17	0.6	4165	2 US-08-477-389-1	Sequence 1, App1
40	17	0.6	4437	1 US-08-559-303B-72	Sequence 72, App1
41	17	0.6	4437	1 US-09-175-828-72	Sequence 72, App1
42	17	0.6	4411529	4 US-09-103-840A-1	Sequence 1, App1
43	16	0.5	36	1 US-08-137-117D-48	Sequence 48, App1
44	16	0.5	36	1 US-08-436-717-48	Sequence 48, App1
45	16	0.5	84	4 US-08-943-731-153	Sequence 153, App
46	16	0.5	132	1 US-08-479-577-1	Sequence 1, App1
47	16	0.5	132	2 US-08-756-977-1	Sequence 1, App1
48	16	0.5	132	5 PCT-US93-05331-5	Sequence 5, App1
49	16	0.5	284	3 US-08-621-018B-20	Sequence 20, App1
50	16	0.5	300	1 US-08-171-385-26	Sequence 26, App1
51	16	0.5	300	3 US-08-361-441B-26	Sequence 26, App1
52	16	0.5	311	4 US-08-916-576B-40	Sequence 40, App1
53	16	0.5	355	4 US-08-991-789A-78	Sequence 78, App1
54	16	0.5	355	4 US-08-991-789A-248	Sequence 248, App
55	16	0.5	444	4 US-08-943-731-127	Sequence 127, App
56	16	0.5	520	3 US-08-899-786-14	Sequence 14, App1
57	16	0.5	562	3 US-08-899-786-16	Sequence 16, App1
58	16	0.5	573	4 US-08-906-156A-67	Sequence 67, App1
59	16	0.5	574	3 US-08-899-786-15	Sequence 15, App1
60	16	0.5	590	4 US-08-906-156A-19	Sequence 19, App1
61	16	0.5	603	4 US-09-385-982-233	Sequence 233, App
62	16	0.5	607	4 US-09-328-111-205	Sequence 205, App
63	16	0.5	625	4 US-09-385-982-101	Sequence 101, App
64	16	0.5	697	1 US-08-468-347-25	Sequence 25, App1
65	16	0.5	697	1 US-08-467-389-25	Sequence 25, App1
66	16	0.5	697	2 US-08-479-779-25	Sequence 25, App1
67	16	0.5	697	2 US-08-469-219-25	Sequence 25, App1
68	16	0.5	697	4 US-09-228-152-25	Sequence 25, App1
69	16	0.5	732	4 US-08-916-576B-5	Sequence 5, App1
70	16	0.5	733	1 US-08-152-485-3	Sequence 3, App1
71	16	0.5	733	1 US-08-463-089-3	Sequence 3, App1
72	16	0.5	733	1 US-08-461-360A-3	Sequence 3, App1
73	16	0.5	733	1 US-08-461-359-3	Sequence 3, App1
74	16	0.5	733	5 PCT-US94-12904-3	Sequence 3, App1
75	16	0.5	795	2 US-08-822-028-11	Sequence 11, App1
76	16	0.5	795	4 US-08-479-285-11	Sequence 11, App1
77	16	0.5	803	3 US-08-454-928-9	Sequence 9, App1
78	16	0.5	843	4 US-09-296-284-24	Sequence 24, App1
79	16	0.5	897	5 PCT-US96-05320A-303	Sequence 303, App
80	16	0.5	921	4 US-09-296-276-7	Sequence 7, App1
81	16	0.5	1047	2 US-08-535-284-3	Sequence 3, App1
82	16	0.5	1054	1 US-08-152-485-1	Sequence 1, App1
83	16	0.5	1054	1 US-08-463-089-1	Sequence 1, App1
84	16	0.5	1054	1 US-08-461-360A-1	Sequence 1, App1
85	16	0.5	1054	1 US-08-461-359-1	Sequence 1, App1
86	16	0.5	1054	5 PCT-US94-12904-1	Sequence 1, App1
87	16	0.5	1081	1 US-08-832-883-57	Sequence 57, App1
88	16	0.5	1081	2 US-08-832-887-57	Sequence 57, App1
89	16	0.5	1103	3 US-08-829-786-13	Sequence 13, App1
90	16	0.5	1104	1 US-08-423-399B-36	Sequence 36, App1
91	16	0.5	1134	4 US-08-432-994A-9	Sequence 9, App1
92	16	0.5	1162	1 US-08-423-399B-34	Sequence 34, App1
93	16	0.5	1166	5 PCT-US96-12129B-1	Sequence 1, App1
94	16	0.5	1173	3 US-08-740-644-1	Sequence 1, App1
95	16	0.5	1178	2 US-08-933-750C-88	Sequence 88, App1
96	16	0.5	1178	2 US-09-234-613-88	Sequence 88, App1
97	16	0.5	1244	2 US-08-204-288-3	Sequence 3, App1
98	16	0.5	1284	1 US-07-882-790-3	Sequence 3, App1
99	16	0.5	1330	2 US-08-868-288A-4	Sequence 4, App1
100	16	0.5	1330	3 US-09-235-373-4	Sequence 4, App1

ALIGNMENTS

RESULT 1
US-08-870-518-34
Sequence 34, Application US/08870518
Patent No. 5925566
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Galcheva-Gargova, Zoya
TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX
TITLE OF INVENTION: PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,518
FILING DATE: 06-JUN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,219
FILING DATE: 06-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/102001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO.: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-870-518-34

Query Match 0.6%; Score 18; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 tttctctcttgatgct 748
|||||

DB 755 ttttctcttgatgct 772

RESULT 2
US-08-747-221B-16/C
Sequence 16, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610e1 Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive

CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: NO. 6063610e1 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ. ID NO.: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1515
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 298
US-08-747-221B-16

Query Match 0.6%; Score 18; DB 3; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
|||||

DB 363 ctccaattcagaatcc 346

RESULT 3
US-08-747-221B-17
Sequence 17, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610e1 Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: NO. 6063610e1 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-747-221B-17

Query Match 0.6%; Score 18; DB 3; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
|||||
DB 1153 CTCCAATTGAGAAATCC 1170

RESULT 4

US-09-005-051-16/c
Sequence 16, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005.051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222e1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1515
FEATURE:

NAME/KEY: Asx = Asn or Asp
LOCATION: 298
US-09-005-051-16

Query Match 0.6%; Score 18; DB 4; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
|||||
DB 363 CTCCAATTGAGAAATCC 346

RESULT 5

US-09-005-051-17
Sequence 17, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005.051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222e1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-005-051-17

Query Match 0.6%; Score 18; DB 4; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
|||||
DB 1153 CTCCAATTGAGAAATCC 1170

RESULT 6
5213972-6/c
Patent No. 5213972

APPLICANT: MCCANDLISS, RUSSELL J.; ANDERSON, DAVID M.
TITLE OF INVENTION: FERMENTATION PROCESS FOR THE PRODUCTION
OF PYRIMIDINE DEOXYRIBONUCLEOSIDES
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/448,158
FILING DATE: 08-DEC-1989
SEQ ID NO: 6:
LENGTH: 1611
5213972-6

Query Match 0.6%; Score 18; DB 6; Length 1611;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 aggattaaaggacttgt 325
|||||

DB 979 AGGATTTAAAGACTTGT 962

RESULT 7

US-08-747-221B-60/c
Sequence 60, Application US/08747221B
Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1650

FEATURE:

NAME/KEY: Asx - Asn or Asp

LOCATION: 433

US-08-747-221B-60

Query Match 0.6%; Score 18; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 46;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 967 ctccaatcagaatcc 984
|||||

DB 498 CTCCAATTCAGAAATCC 481

RESULT 8

US-08-747-221B-61
Sequence 61, Application US/08747221B
Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-747-221B-61

Query Match 0.6%; Score 18; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 967 ctccaatcagaatcc 984
|||||

DB 1153 CTCCAATTCAGAAATCC 1170

RESULT 9

US-09-005-051-60/c
Sequence 60, Application US/09005051
Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

```
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005.051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: NO. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1650
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 433
US-09-005-051-60

Query Match 0.6%; Score 18; DB 4; Length 1650;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 967 ctccaattcgaatcc 984
Db 498 CTCCAATTGAGAAATCC 481

RESULT 10
US-09-005-051-61
Sequence 61, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/09/005.051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: NO. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-005-051-61

Query Match 0.6%; Score 18; DB 4; Length 1650;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 967 ctccaattcgaatcc 984
Db 1153 CTCCAATTGAGAAATCC 1170

RESULT 11
US-08-857-213-2/c
Sequence 2, Application US/08857213
Patent No. 6054290
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Murty, Lynn E.
TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/857,213
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0297 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 base pairs
TYPE: nucleic acid
```

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBRNGT01
CLONE: 148415
US-08-857-213-2

Query Match 0.6%; Score 18; DB 3; Length 1721;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaattcataagac 18
|||||
DB 1538 ATGAATTCATAAAGAC 1521

RESULT 12
US-08-139-937-11
Sequence 11, Application US/08139937
Patent No. 5821070
GENERAL INFORMATION:
APPLICANT: LEE, MEN-HWA
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: CELLULAR GENES ENCODING
RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/139,937
FILING DATE: 20-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/979,156
FILING DATE: 20-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9370
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-139-937-11

Query Match 0.6%; Score 18; DB 1; Length 1800;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2796 aagaaccctctgaaaa 2813
|||||
DB 307 AAGAACCCTCTGAAAA 324

RESULT 13

PCT-US93-11310-11
Sequence 11, Application PC/TUS9311310
GENERAL INFORMATION:
APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
TITLE OF INVENTION: CELLULAR GENES ENCODING
RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11310
FILING DATE: 19-NOV-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9790
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US93-11310-11

Query Match 0.6%; Score 18; DB 5; Length 1800;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2796 aagaaccctctgaaaa 2813
|||||
DB 307 AAGAACCCTCTGAAAA 324

RESULT 14
US-08-747-221B-13/C
Sequence 13, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
Molecules, Proteins and Uses Thereof
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/484-9505
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1517
FEATURE:
NAME/KEY: Asx - Asn or Asp
LOCATION: 300
US-08-747-221B-13

Query Match 0.6%; Score 18; DB 3; Length 1982;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 967 ctccaattcagaatcc 984
|||||

Db 365 CTCGAATTGAGAAATCC 348

RESULT 15
US-08-747-221B-15
Sequence 15, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/484-9505
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-747-221B-15

Query Match 0.6%; Score 18; DB 3; Length 1982;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 967 ctccaattcagaatcc 984
|||||

Db 1618 CTCGAATTGAGAAATCC 1635

RESULT 16
US-09-005-051-13/c
Sequence 13, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/484-9505
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1517
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 300
US-09-005-051-13

Query Match 0.6%; Score 18; DB 4; Length 1982;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 967 ctccaattcagaatcc 984

Db 365 CTCGAATTCAGAAATCC 348

RESULT 17

US-09-005-051-15
Sequence 15, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ. ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-005-051-15

Query Match 0.6%; Score 18; DB 4; Length 1982;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
Db 1618 CTCGAATTCAGAAATCC 1635

RESULT 18

US-08-747-221B-57/c
Sequence 57, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation

STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ. ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 2144 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1682
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 462
US-08-747-221B-57

Query Match 0.6%; Score 18; DB 3; Length 2144;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
Db 527 CTCGAATTCAGAAATCC 510

RESULT 19

US-08-747-221B-59
Sequence 59, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Verser, Carol Talkington
: REGISTRATION NUMBER: 37,459
: REFERENCE/DOCKET NUMBER: FC-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 970/493-7272
: TELEFAX: 970/484-9505
: INFORMATION FOR SEQ ID NO: 59:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2144 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-08-747-221B-59

Query Match
Best Local Similarity 100.0%; Score 18; DB 3; Length 2144;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 967 ctccaattcagaattcc 984
Db 1618 CTCCAATTGAGAAATCC 1635

RESULT 20
US-09-005-051-57/C
: Sequence 57, Application US/09005051
: Patent No. 6291222
: GENERAL INFORMATION:
: APPLICANT: Silver, Gary W.
: APPLICANT: Wisniewski, Nancy
: TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
: TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
: NUMBER OF SEQUENCES: 66
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Heska Corporation Verser, Ph.D.
: STREET: 1825 Sharp Point Drive
: CITY: Fort Collins
: STATE: Colorado
: COUNTRY: USA
: ZIP: 80525
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Windows 95
: SOFTWARE: WordPerfect for Windows, Version 7.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/005,051
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/747,221
: FILING DATE: No. 6291222ember 12, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verser, Carol Talkington
: REGISTRATION NUMBER: 37,459
: REFERENCE/DOCKET NUMBER: FC-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 970/493-7272
: TELEFAX: 970/484-9505
: INFORMATION FOR SEQ ID NO: 57:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2144 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 30..1682
```

```

: FEATURE:
: NAME/KEY: Asx = Asn or Asp
: LOCATION: 462
: US-09-005-051-57

Query Match
Best Local Similarity 100.0%; Score 18; DB 4; Length 2144;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 967 ctccaattcagaattcc 984
Db 527 CTCCAATTGAGAAATCC 510

RESULT 21
US-09-005-051-59
: Sequence 59, Application US/09005051
: Patent No. 6291222
: GENERAL INFORMATION:
: APPLICANT: Silver, Gary W.
: APPLICANT: Wisniewski, Nancy
: TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
: TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
: NUMBER OF SEQUENCES: 66
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Heska Corporation Verser, Ph.D.
: STREET: 1825 Sharp Point Drive
: CITY: Fort Collins
: STATE: Colorado
: COUNTRY: USA
: ZIP: 80525
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Windows 95
: SOFTWARE: WordPerfect for Windows, Version 7.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/005,051
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/747,221
: FILING DATE: No. 6291222ember 12, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verser, Carol Talkington
: REGISTRATION NUMBER: 37,459
: REFERENCE/DOCKET NUMBER: FC-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 970/493-7272
: TELEFAX: 970/484-9505
: INFORMATION FOR SEQ ID NO: 59:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2144 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-09-005-051-59

Query Match
Best Local Similarity 100.0%; Score 18; DB 4; Length 2144;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 967 ctccaattcagaattcc 984
Db 1618 CTCCAATTGAGAAATCC 1635

RESULT 22
US-09-192-104-1
```

```
; Sequence 1, Application US/09192104B
; Patent No. 6184020
; GENERAL INFORMATION:
; APPLICANT: Alexander Blinkovsky
; APPLICANT: Tony Byun
; APPLICANT: Alan V. Klotz
; APPLICANT: Alan Sloma
; APPLICANT: Maria Tang
; APPLICANT: Mikio Fujii
; APPLICANT: Chigusa Marumoto
; APPLICANT: Lene Venke Kofod
; TITLE OF INVENTION: Polypeptides Having Aminoepitidase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5379.200-US
; CURRENT APPLICATION NUMBER: US/09/192,104B
; EARLIER FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/069719
; EARLIER FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: 1465/97
; EARLIER FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: PA 1998 00670
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Sphingomonas
US-09-192-104-1
```

```
Query Match      0.6%; Score 18; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2092 gctggtgtgctggaagc 2109
      ||||||||||||||||
Db 1911 gctggtgtgctggaagc 1928
```

```
RESULT 23
US-09-543-446-1
; Sequence 1, Application US/09543446
; Patent No. 6303360
; GENERAL INFORMATION:
; APPLICANT: Alexander Blinkovsky
; APPLICANT: Tony Byun
; APPLICANT: Alan V. Klotz
; APPLICANT: Alan Sloma
; APPLICANT: Maria Tang
; APPLICANT: Mikio Fujii
; APPLICANT: Chigusa Marumoto
; APPLICANT: Lene Venke Kofod
; TITLE OF INVENTION: Polypeptides Having Aminoepitidase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5379.210-US
; CURRENT APPLICATION NUMBER: US/09/543,446
; EARLIER FILING DATE: 2000-04-05
; EARLIER APPLICATION NUMBER: 60/069719
; EARLIER FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: 1465/97
; EARLIER FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: PA 1998 00670
; EARLIER FILING DATE: 1998-05-15
; EARLIER APPLICATION NUMBER: 09/192,104
; EARLIER FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Sphingomonas
US-09-543-446-1
```

```
Query Match      0.6%; Score 18; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2092 gctggtgtgctggaagc 2109
      ||||||||||||||||
Db 1911 gctggtgtgctggaagc 1928
```

```
RESULT 24
US-08-699-103B-1/c
; Sequence 1, Application US/08699103B
; Patent No. 6107462
; GENERAL INFORMATION:
; APPLICANT: Rine, Jasper D.
; APPLICANT: Hampton, Randolph
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/699,103B
; FILING DATE: 16-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/002,581
; FILING DATE: 17-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 09272/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/322-5070
; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-699-103B-1
```

```
Query Match      0.6%; Score 18; DB 3; Length 4982;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 356 cccttggtgaagataty 373
      ||||||||||||||||
Db 1743 CCCTTGCTGAAGATATGT 1726
```

```
RESULT 25
US-09-037-990B-27
; Sequence 27, Application US/09037990B
; Patent No. 6248519
; GENERAL INFORMATION:
; APPLICANT: ENGEL, Stacia R.
; DESCENZO, Richard A.
; MORENZONI, Richard A.
US-09-037-990B-27
```



```

; IRELAN, Nancy A.
; TITLE OF INVENTION: DETECTION OF FERMENTATION-RELATED
; MICROORGANISMS
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,990B
; FILING DATE: 11-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Jeffrey S.
; REGISTRATION NUMBER: 31,879
; REFERENCE/DOCKET NUMBER: 29520/30001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SQUENCE DESCRIPTION: SEQ ID NO: 27:
;
; US-09-037-990B-27
;
; Query Match      0.6%; Score 17; DB 4; Length 290;
; Best Local Similarity 100.0%; Pred. No. 1.4e+02;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Oy 1117 aaaaacaacacaaca 1133
;      |||||||||||||||
;      209 AAAAACAACACAACA 225
;
; RESULT 26
; US-08-991-789A-171/c
; Sequence 171, Application US/08991789A
; Patent No. 6225054
;
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
;              Smith, John M.
;              Reed, Steven G.
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
;
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; IRELAN, Nancy A.
; TITLE OF INVENTION: DETECTION OF FERMENTATION-RELATED
; MICROORGANISMS
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,990B
; FILING DATE: 11-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Jeffrey S.
; REGISTRATION NUMBER: 31,879
; REFERENCE/DOCKET NUMBER: 29520/30001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661 base pairs
```

```

; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potler, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 171:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 171:
;
; US-08-991-789A-171
;
; Query Match      0.6%; Score 17; DB 4; Length 383;
; Best Local Similarity 100.0%; Pred. No. 1.4e+02;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Oy 374 acattattttaacttg 390
;      |||||||||||||||
;      Db 35 ACATTATTTTTAACTTG 19
;
; RESULT 27
; US-09-037-990B-8
; Sequence 8, Application US/09037990B
; Patent No. 6248519
;
; GENERAL INFORMATION:
; APPLICANT: ENGEL, Stacia R.
;              DESCENZO, Richard A.
;              MORENZONI, Richard A.
;              IRELAN, Nancy A.
;
; TITLE OF INVENTION: DETECTION OF FERMENTATION-RELATED
; MICROORGANISMS
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,990B
; FILING DATE: 11-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Jeffrey S.
; REGISTRATION NUMBER: 31,879
; REFERENCE/DOCKET NUMBER: 29520/30001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661 base pairs
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414)277-5591
INFORMATION FOR SEQ ID NO.: 183:
SEQUENCE CHARACTERISTICS:
LENGTH: 1563 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Klebsiella pneumoniae
US-08-743-637B-183

Query Match 0.6%; Score 17; DB 2; Length 1563;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 470 tccctcagcctctcag 486
|||||
Db 74 TCCTCAGGCTTTCAG 58

RESULT 31
US-09-318-443-7/C
Sequence 7, Application US/09318443
Patent No. 6197947
GENERAL INFORMATION:
APPLICANT: Hemmati-Briyanlou, Ali
APPLICANT: Weinstein, Daniel C.
TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
TITLE OF INVENTION: THEREOF
FILE REFERENCE: 600-1-211 N
CURRENT APPLICATION NUMBER: US/09/318,443
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1682
TYPE: DNA
ORGANISM: Homo sapiens
US-09-318-443-7

Query Match 0.6%; Score 17; DB 4; Length 1682;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1339 caggagttacacagcagg 1355
|||||
Db 1501 CAGGAGTACACAGG 1485

RESULT 32
US-09-178-252-26
Sequence 26, Application US/09178252
Patent No. 6218188
GENERAL INFORMATION:
APPLICANT: Cardineau, Guy A.
APPLICANT: Steiman, Steven J.
APPLICANT: Narva, Kenneth E.
TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins
FILE REFERENCE: MA-714XC2
CURRENT APPLICATION NUMBER: US/09/178,252
CURRENT FILING DATE: 1998-10-23
EARLIER APPLICATION NUMBER: 60/065,215
EARLIER FILING DATE: 1997-11-12
EARLIER APPLICATION NUMBER: 60/076,445
EARLIER FILING DATE: 1998-03-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26

LENGTH: 1965
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic B.t. toxin gene
US-09-178-252-26

Query Match 0.6%; Score 17; DB 4; Length 1965;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 428 accaacaccatcacgc 444
|||||
Db 1519 accaacaccatcacgc 1535

RESULT 33
US-08-960-780-10
Sequence 10, Application US/08960780
Patent No. 6204435
GENERAL INFORMATION:
APPLICANT: Felleison, Jerald S.
APPLICANT: Schepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeltz, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanichik, Lloyd & Saliwanichik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanichik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 158C2-p11
US-08-960-780-10

Query Match 0.6%; Score 17; DB 4; Length 2035;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2585 aagatggaatgaagct 2601
|||||
DB 1973 AAGATGGAATGAAGCT 1989

RESULT 34
US-09-073-898-10
; Sequence 10, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Feltelson, Jerald S.
; APPLICANT: Schmepl, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Flustad-lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 158C2-pt1
; US-09-073-898-10

Query Match 0.6%; Score 17; DB 4; Length 2035;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2585 aagatggaatgaagct 2601

|||||
DB 1973 AAGATGGAATGAAGCT 1989

RESULT 35
US-09-115-446-1/C
; Sequence 1, Application US/09115446
; Patent No. 6165719
; GENERAL INFORMATION:
; APPLICANT: Chandry, George K.
; APPLICANT: Gargus, Jay J.
; APPLICANT: Gutman, George
; APPLICANT: Fanti, Emmanuelle
; APPLICANT: Kaiman, Katarin
; TITLE OF INVENTION: hKCA3/KCN3 SMALL CONDUCTANCE CALCIUM
; TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
; TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
; FILE REFERENCE: 07306/014001
; CURRENT APPLICATION NUMBER: US/09/115,446
; CURRENT FILING DATE: 1998-07-14
; EARLIER APPLICATION NUMBER: 60/052,556
; EARLIER FILING DATE: 1997-07-15
; EARLIER APPLICATION NUMBER: 60/070,741
; EARLIER FILING DATE: 1998-01-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2521
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (287)...(2479)
; US-09-115-446-1

Query Match 0.6%; Score 17; DB 4; Length 2521;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 tgctgctgaagctggcg 720
|||||
DB 492 TGCTGCTGAAGCTGGCG 476

RESULT 36
US-09-115-446-5/C
; Sequence 5, Application US/09115446
; Patent No. 6165719
; GENERAL INFORMATION:
; APPLICANT: Chandry, George K.
; APPLICANT: Gargus, Jay J.
; APPLICANT: Gutman, George
; APPLICANT: Fanti, Emmanuelle
; APPLICANT: Kaiman, Katarin
; TITLE OF INVENTION: hKCA3/KCN3 SMALL CONDUCTANCE CALCIUM
; TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
; TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
; FILE REFERENCE: 07306/014001
; CURRENT APPLICATION NUMBER: US/09/115,446
; CURRENT FILING DATE: 1998-07-14
; EARLIER APPLICATION NUMBER: 60/052,556
; EARLIER FILING DATE: 1997-07-15
; EARLIER APPLICATION NUMBER: 60/070,741
; EARLIER FILING DATE: 1998-01-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2526
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-115-446-5

Query Match 0.6%; Score 17; DB 4; Length 2526;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 704 tgcctcgaagctgcg 720
|||||
Db 497 tgcctcgaagctgcg 481

RESULT 37

US-08-095-737-1/c
; Sequence 1, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095.737
; FILING DATE: 19930722
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A
; REGISTRATION NUMBER: 29.655
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..2709
US-08-095-737-1

Query Match 0.6%; Score 17; DB 1; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2776 atttagtgcatlctt 2792
|||||
Db 4070 atttagtgcatlctt 4054

RESULT 38

US-08-480-145-1/c
; Sequence 1, Application US/08480145
; Patent No. 5717067
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P

APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.145
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095.737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A
; REGISTRATION NUMBER: 29.655
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..2709
US-08-480-145-1

Query Match 0.6%; Score 17; DB 1; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2776 atttagtgcatlctt 2792
|||||
Db 4070 atttagtgcatlctt 4054

RESULT 39

US-08-477-389-1/c
; Sequence 1, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 21..2709
US-08-477-389-1

Query Match 0.6%; Score 17; DB 2; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2776 attagatgatattt 2792
|||||
Db 4070 ATTTAGTGATTTT 4054

RESULT 40
US-08-559-303B-72
Sequence 72, Application US/08559303B
Patent No. 5824501
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-559-303B-72

Query Match 0.6%; Score 17; DB 1; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3017 aattgatgatgat 3033
|||||
Db 943 AATTGATGATGAT 959

RESULT 41
US-09-175-828-72
Sequence 72, Application US/09175828
Patent No. 6221643
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175,828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES

ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-175-828-72

Query Match 0.6%; Score 17; DB 4; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3017 aattgatgatgatgat 3033
|||||

DB 943 AATTGATGATGATGAT 959

RESULT 42
US-09-103-840A-1/C
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 0.6%; Score 17; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2365 aacctgaagaagatgtg 2381
|||||

DB 3725573 AACCTGAGAGATGTG 3725557

RESULT 43
US-08-137-117D-48/C
Sequence 48, Application US/08137117D
Patent No. 5795965
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: BENDIG, Mary
APPLICANT: JONES, Steven
APPLICANT: SALDANHA, Jose
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
NUMBER OF SEQUENCES: 158
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/137,117D
FILING DATE: 20-DEC-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP92/00544
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-32084
FILING DATE: 19-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-95476
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/126/AAOK
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-137-117D-48

Query Match 0.5%; Score 16; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2147 ctctcatgtgtgaagc 2162
|||||

DB 22 CTCATGATGTGAAGC 7

RESULT 44
US-08-436-717-48/C
Sequence 48, Application US/08436717
Patent No. 5817790
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: BENDIG, Mary
APPLICANT: JONES, Steven
APPLICANT: SALDANHA, Jose
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
NUMBER OF SEQUENCES: 158
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,717
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/137,117

;; FILING DATE: 20-DEC-1993
;; APPLICATION NUMBER: WO PCT/JP92/00544
;; FILING DATE: 24-APR-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 4-32084
;; FILING DATE: 19-FEB-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 3-95476
;; FILING DATE: 25-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: WEGNER, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 53466/126/AAOK
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 48:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 36 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-436-717-48

Query Match 0.5%; Score 16; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2147 cctcatgtgtgaagc 2162
DB 22 cctcatgtgtgaagc 7

;; RESULT 45
;; US-08-943-731-153/C
;; Sequence 153, Application US/08943731
;; Patent No. 6265157
;; GENERAL INFORMATION:
;; APPLICANT: PROCKOP, DARWIN J.
;; APPLICANT: SPOTILIA, LORETTA D.
;; APPLICANT: DELITAS, CONSTANTINOS D.
;; APPLICANT: SEREDA, LARISA
;; APPLICANT: LARSON, ANDREA W.
;; APPLICANT: PACK, MICHAEL
;; APPLICANT: COLIGE, ALAIN
;; APPLICANT: EARLY, JAMES
;; APPLICANT: KORKKO, JARMO
;; APPLICANT: ALA-KORKKO, LEBENA, et al.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;; NUMBER OF SEQUENCES: 666
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
;; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;; STREET: FLR.
;; CITY: PHILADELPHIA
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103-7086
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/943,731
;; FILING DATE: 03-OCT-1997
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/212,322
;; FILING DATE: 14-MAR-1994

;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 07/803,628
;; FILING DATE: 03-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DOYLE LEARY Ph.D., KATHRYN
;; REGISTRATION NUMBER: 36,317
;; REFERENCE/DOCKET NUMBER: 9598-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-965-1284
;; TELEFAX: 215-567-2991
;; TELEX: 831-494
;; INFORMATION FOR SEQ ID NO: 153:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 84 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-943-731-153

Query Match 0.5%; Score 16; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1384 catgagccagagag 1399
DB 61 catgagccagagag 46

;; RESULT 46
;; US-08-479-577-1/C
;; Sequence 1, Application US/08479577
;; Patent No. 5633449
;; GENERAL INFORMATION:
;; APPLICANT: Milton Zaitlin, and Peter Palukaitis
;; TITLE OF INVENTION: Induction Of Resistance To.
;; TITLE OF INVENTION: Viral Diseases In Plants
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Yahwak & Associates
;; STREET: 25 Skytop Drive
;; CITY: Trumbull
;; STATE: Connecticut
;; COUNTRY: USA
;; ZIP: 06611
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: Macintosh
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Microsoft Word 4.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/479,577
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US/08/197,096
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: George M. Yahwak
;; REGISTRATION NUMBER: 26,824
;; REFERENCE/DOCKET NUMBER: CRF D-1123NM
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (203)268-1951
;; TELEFAX: (203)268-1951
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 132 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: RNA
;; US-08-479-577-1

Query Match 0.5%: Score 16; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2506 gcaatgaaatcctag 2521
|||||
Db 100 GCACTGAAATCCTAG 85

RESULT 47

US-08-756-977-1/c
; Sequence 1, Application US/08756977
; Patent No. 5945581
; GENERAL INFORMATION:
; APPLICANT: Zaitlin, Milton
; APPLICANT: Palukaitis, Peter
; TITLE OF INVENTION: INDUCTION OF RESISTANCE TO VIRAL
; TITLE OF INVENTION: DISEASES IN PLANTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,977
; FILING DATE: 26-NOV-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (genomic)
; US-08-756-977-1

Query Match 0.5%: Score 16; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2506 gcaatgaaatcctag 2521
|||||
Db 100 GCACTGAAATCCTAG 85

RESULT 48
PCT-US93-05331-5/c
; Sequence 5, Application PC/TUS9305331
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Induction Of Resistance To
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates

STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05331
FILING DATE: 19930603
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-1123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

PCT-US93-05331-5

Query Match 0.5%: Score 16; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2506 gcaatgaaatcctag 2521
|||||
Db 100 GCACTGAAATCCTAG 85

RESULT 49

US-08-621-018B-20
; Sequence 20, Application US/08621018B
; Patent No. 6060239
; GENERAL INFORMATION:
; APPLICANT: Stuart, Susan G.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Selthamer, Jeffrey J.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: CELLULBREVIN HOMOLOGS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,018B
; FILING DATE: March 22, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/409,373
; FILING DATE: March 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0029-1 CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: SYNORAT01
CLONE: 370165
US-08-621-018B-20

Query Match 0.5%; Score 16; DB 3; Length 284;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2151 catggtggaagccagt 2166
Db 43 CATGCTGAAGCCAGT 58

RESULT 50
US-08-171-385-26
Sequence 26, Application US/08171385
Patent No. 5527884
GENERAL INFORMATION:
APPLICANT: Mary E. Russell
APPLICANT: Ulrike Utans
TITLE OF INVENTION: Mediators of Chronic Allergraft
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171,385
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/006001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 300
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-171-385-26

Query Match 0.5%; Score 16; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1960 aactggaagcaggaat 1975
Db 282 AACTGGAAGCAGGAAT 297

RESULT 51
US-08-361-441B-26
Sequence 26, Application US/08361441B
Patent No. 6077948
GENERAL INFORMATION:
APPLICANT: Russell, Mary E.
APPLICANT: Utans, Ulrike
TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/361,441B
FILING DATE: 21-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/171,385
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/014001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-361-441B-26

Query Match 0.5%; Score 16; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1960 aactggaagcaggaat 1975
Db 282 AACTGGAAGCAGGAAT 297

RESULT 52
US-08-916-576B-40
Sequence 40, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DITTON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-916-576B-40

Query Match 0.5%; Score 16; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1785 attgactcttga 1800
|||||
DB 108 attgactcttga 123

RESULT 53
US-08-991-789A-78
Sequence 78, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-08-991-789A-78

Query Match 0.5%; Score 16; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2522 ctcaaatcttcaaa 2537
|||||
DB 149 ctcaaatcttcaaa 164

RESULT 54
US-08-991-789A-248
Sequence 248, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 248:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 248:
US-08-991-789A-248

Query Match 0.5%; Score 16; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2522 ctcaaatcttcaaa 2537
|||||

Db 149 CTCGAGATCTTCACAA 164

RESULT 55

US-08-943-731-127

Sequence 127, Application US/08943731

Patent No. 6265157

GENERAL INFORMATION:

APPLICANT: PROCKOP, DARWIN J.

APPLICANT: SPOTILA, LORETTA D.

APPLICANT: DELTAS, CONSTANTINOS D.

APPLICANT: SEREDA, LARISA

APPLICANT: LARSON, ANDREA W.

APPLICANT: PACK, MICHAEL

APPLICANT: COLIGE, ALAIN

APPLICANT: EARLY, JAMES

APPLICANT: KORKKO, JARMO

APPLICANT: ALA-KOKKO, LEENA, et al.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING

TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES

NUMBER OF SEQUENCES: 666

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.

STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND

CITY: PHILADELPHIA

STATE: PA

COUNTRY: USA

ZIP: 19103-7086

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/943,731

FILING DATE: 03-OCT-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,322

FILING DATE: 14-MAR-1994

APPLICATION DATA:

APPLICATION NUMBER: US 07/803,628

FILING DATE: 03-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: DOYLE LEARY Ph.D., KATHRYN

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9598-27

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-965-1284

TELEFAX: 215-567-2991

TELEX: 831-494

INFORMATION FOR SEQ ID NO: 127:

SEQUENCE CHARACTERISTICS:

LENGTH: 434 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-943-731-127

Query Match

Best Local Similarity 0.5%; Score 16; DB 4; Length 434;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 196 TGATCTTGTCTCTC 211

RESULT 56

US-08-899-786-14/c

Sequence 14, Application US/08899786

Patent No. 6001572

GENERAL INFORMATION:

APPLICANT: Toothman, Penelope

TITLE OF INVENTION: Method of Identifying Aloe Using

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Avenue, Suite 200

CITY: Englewood

STATE: Colorado

COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB

COMPUTER: IBM pc compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,786

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/022,611

FILING DATE: 26-JULY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: UNI.07

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3433

TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 520 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

US-08-899-786-14

Query Match

Best Local Similarity 0.5%; Score 16; DB 3; Length 520;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 98 GATTGGCTCAGATT 83

RESULT 57

US-08-899-786-16/c

Sequence 16, Application US/08899786

Patent No. 6001572

GENERAL INFORMATION:

APPLICANT: Toothman, Penelope

TITLE OF INVENTION: Method of Identifying Aloe Using

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Avenue, Suite 200

CITY: Englewood

STATE: Colorado

COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB

COMPUTER: IBM pc compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,786

FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/022,611
FILING DATE: 26-JULY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: UNI.07
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 562 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-899-786-16

Query Match 0.5%; Score 16; DB 3; Length 562;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 gatttgctcagatt 313
|||||
DB 100 GATTGGCTCAGATT 85

RESULT 58

US-08-906-156A-67/c
Sequence 67, Application US/08906156A
Patent No. 6287854
GENERAL INFORMATION:
APPLICANT: SPURR, NIGEL K
APPLICANT: GRAY, IAN C
APPLICANT: STEWART, LORNA M
TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,156A
FILING DATE: 05-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,655
FILING DATE: 02-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,147
FILING DATE: 13-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,840
FILING DATE: 23-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/96GB/02588
FILING DATE: 22-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 1090-14
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: human
US-08-906-156A-67

Query Match 0.5%; Score 16; DB 4; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 tcatcacaatgatctt 176
|||||
DB 76 TCATTCACATGATTTT 61

RESULT 59

US-08-899-786-15/c
Sequence 15, Application US/08899786
Patent No. 6001572
GENERAL INFORMATION:
APPLICANT: Toothman, Penelope
TITLE OF INVENTION: Method of Identifying Aloe Using
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,786
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/022,611
FILING DATE: 26-JULY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: UNI.07
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-899-786-15

Query Match 0.5%; Score 16; DB 3; Length 574;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 gattggctcagatt 313
|||||
DB 112 GATTGGCTCAGATT 97

RESULT 60

US-08-906-156A-19/c
; Sequence 19, Application US/08906156A
; Patent No. 6287854
; GENERAL INFORMATION:
; APPLICANT: SPURR, NIGEL K
; APPLICANT: GRAY, IAN C
; APPLICANT: STEWART, LORNA M
; TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER
; TITLE OF INVENTION: AND TREATMENT THEREOF
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYTE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,156A
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,655
; FILING DATE: 02-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,147
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,840
; FILING DATE: 23-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/96GB/02588
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1090-14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Part of gene corresponding to IMAGE 264611
US-08-906-156A-19

Query Match 0.5%; Score 16; DB 4; Length 590;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 tcattcacatgatttt 176
|||||
DB 76 TCATTCACTCATGATT 61

RESULT 61
US-09-385-982-233
; Sequence 233, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 233
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(603)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-233

Query Match 0.5%; Score 16; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2377 atgtgttatttcatt 2392
|||||
DB 373 atgtgttatttcatt 388

RESULT 62
US-09-328-111-205/c
; Sequence 205, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O.
; APPLICANT: STELMANN, KATHLEEN E.
; APPLICANT: ASTLE, JON H.
; APPLICANT: BURGESS, CHRISTOPHER C.
; APPLICANT: BUSHNELL, STEVEN E.
; APPLICANT: CARROLL III, EDDIE
; APPLICANT: CATINO, THEODORE J.
; APPLICANT: DERTI, ADAM
; APPLICANT: FORD, DONNA M.
; APPLICANT: LEWIS, MARCIA E.
; APPLICANT: MONAHAN, JOHN E.
; APPLICANT: SCHLEGEL, ROBERT
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 205
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(607)

OTHER INFORMATION: n = A,T,C or G
US-09-328-111-205

Query Match 0.5%; Score 16; DB 4; Length 607;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2942 cagcattagtcagaaa 2957
|||||

Db 325 CAGCATTAGTCAGAAA 310

RESULT 63

US-09-385-982-101/C

Sequence 101, Application US/09385982

Patent No. 6262334

GENERAL INFORMATION:

APPLICANT: ENDEGE, WILSON O., ET AL.

TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

FILE REFERENCE: CCDDA-260XX

CURRENT APPLICATION NUMBER: US/09/385,982

EARLIER FILING DATE: 1999-08-30

EARLIER APPLICATION NUMBER: 09/328,111

EARLIER FILING DATE: 1999-06-08

EARLIER APPLICATION NUMBER: 60/117,393

EARLIER FILING DATE: 1999-01-27

EARLIER APPLICATION NUMBER: 60/098,639

NUMBER OF SEQ ID NOS: 544

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 101

LENGTH: 625

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(625)

OTHER INFORMATION: n = A,T,C or G

US-09-385-982-101

Query Match 0.5%; Score 16; DB 4; Length 625;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 728 ttcttctctctctga 743
|||||

Db 159 TTCCTTCTCTCTGA 144

RESULT 64

US-08-468-347-25/C

Sequence 25, Application US/08468347

Patent No. 5783421

GENERAL INFORMATION:

APPLICANT: Zeelon, Elisha P.

APPLICANT: Weider, Moshe M.

APPLICANT: Levanon, Avigdor

TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,347

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/225,442

FILING DATE: 08-APR-1994

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-977-9550

TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 697 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

US-08-468-347-25

Query Match 0.5%; Score 16; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2905 tttttgacttagta 2920
|||||

Db 678 TTTTGTGACTTAGTA 663

RESULT 65

US-08-467-389-25/C

Sequence 25, Application US/08467389

Patent No. 5824641

GENERAL INFORMATION:

APPLICANT: Zeelon, Elisha P.

APPLICANT: Weider, Moshe M.

APPLICANT: Levanon, Avigdor

TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,389

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/225,442

FILING DATE: 08-APR-1994

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-977-9550

TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 25;
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-467-389-25

Query Match 0.5%; Score 16; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2905 ttlttgacttagta 2920
|||||
Db 678 TTTTTCACCTTAGTA 663

RESULT 66

US-08-779-379-25/C
Sequence 25, Application US/08779379
Patent No. 5858970
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
TITLE OF INVENTION: INHIBITORY ACTIVITY
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/779,379
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-779-379-25

Query *Match 0.5%; Score 16; DB 2; Length 697;

Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2905 ttlttgacttagta 2920
|||||
Db 678 TTTTTCACCTTAGTA 663

RESULT 67

US-08-469-219-25/C
Sequence 25, Application US/08469219
Patent No. 5863534
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
TITLE OF INVENTION: INHIBITORY ACTIVITY
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,219
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-469-219-25

Query Match 0.5%; Score 16; DB 2; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2905 ttlttgacttagta 2920
|||||
Db 678 TTTTTCACCTTAGTA 663

RESULT 68

US-09-228-152-25/C
Sequence 25, Application US/09228152
Patent No. 6211341
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.

APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa INHIBITORY ACTIVITY
FILE REFERENCE: 430208YA
CURRENT APPLICATION NUMBER: US/09/228,152
CURRENT FILING DATE: 1999-01-11
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 25
LENGTH: 697
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
OTHER INFORMATION: of cDNA clone psp65-xai-13.
US-09-228-152-25

Query Match 0.5%; Score 16; DB 4; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2905 tttttgacttagta 2920
|||||
Db 678 tttttgacttagta 663

RESULT 69
US-08-916-576B-5/C
Sequence 5, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488, 0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 49..546

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 116..546
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 49..117
US-08-916-576B-5

Query Match 0.5%; Score 16; DB 4; Length 732;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1785 atttgactttgaa 1800
|||||
Db 589 atttgacttttgaa 574

RESULT 70
US-08-152-485-3
Sequence 3, Application US/08152485
Patent No. 5539094
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Takayama, Shinichi
APPLICANT: Sato, Takaaki
TITLE OF INVENTION: Bcl-2-associated proteins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/152,485
FILING DATE: 10-NOV-1993
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9725
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-152-485-3

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2697 gtgaacatttgag 2712
|||||
Db 229 gtgaacatttgag 244

RESULT 71
US-08-463-089-3
Sequence 3, Application US/08463089

```
Patent No. 5641866
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Takayama, Shinichi
APPLICANT: Sato, Takaaki
TITLE OF INVENTION: Bcl-2-associated proteins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,089
FILING DATE: 10-NOV-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9725
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-463-089-3

Query Match      0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2697 gtgaaacattgag 2712
Db 229 GTGAACATTGAG 244

RESULT 72
US-08-461-360A-3
Sequence 3, Application US/08461360A
Patent No. 5650491
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Takayama, Shinichi
APPLICANT: Sato, Takaaki
TITLE OF INVENTION: Bcl-2-associated proteins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,360A
```

```
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9725
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-461-360A-3

Query Match      0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2697 gtgaaacattgag 2712
Db 229 GTGAACATTGAG 244

RESULT 73
US-08-461-359-3
Sequence 3, Application US/08461359
Patent No. 5686595
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Takayama, Shinichi
APPLICANT: Sato, Takaaki
TITLE OF INVENTION: Bcl-2-associated proteins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,359
FILING DATE: 10-NOV-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9725
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 733 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-461-359-3

Query Match      0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2697 gtgaacattggag 2712
 |||
 Db 229 GTGAACATTGGAG 244

RESULT 74
 PCT-US94-12904-3
 : Sequence 3, Application PC/TUS9412904
 : GENERAL INFORMATION:
 : APPLICANT: Reed, John C.
 : APPLICANT: Takayama, Shinichi
 : APPLICANT: Sato, Takaaki
 : TITLE OF INVENTION: Bcl-2-associated proteins
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: CAMPBELL AND FLORES
 : STREET: 4370 La Jolla Village Drive, Suite 700
 : CITY: San Diego
 : STATE: California
 : COUNTRY: USA
 : ZIP: 92122
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US94/12904
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/152,485
 : FILING DATE: 12-NOV-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Campbell, Cathryn A.
 : REGISTRATION NUMBER: 31,815
 : REFERENCE/DOCKET NUMBER: PP-LJ 1201
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (619) 535-9001
 : TELEFAX: (619) 535-8949
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 733 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : PCT-US94-12904-3

Query Match 0.5%; Score 16; DB 5; Length 733;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2697 gtgaacattggag 2712
 |||
 Db 229 GTGAACATTGGAG 244

RESULT 75
 US-08-822-028-11
 : Sequence 11, Application US/08822028
 : Patent No. 5993813
 : GENERAL INFORMATION:
 : APPLICANT: MEZES, PETER S
 : APPLICANT: GOURLIE, BRIAN B
 : APPLICANT: RIXON, MARK W
 : APPLICANT: ANDERSON, WH KERR
 : APPLICANT: KAPLAN, DONALD A
 : APPLICANT: SCHOLIM, JEFFREY
 : TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,

: TITLE OF INVENTION: MODIFIED ANTIBODIES FOR CANCER TREATMENT
 : NUMBER OF SEQUENCES: 74
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: DUANE C ULMER
 : STREET: P.O. BOX 1967
 : CITY: MIDLAND
 : STATE: MICHIGAN
 : COUNTRY: USA
 : ZIP: 48641-1967
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/822,028
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/040,687
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: ULMER, DUANE C
 : REGISTRATION NUMBER: 34,941
 : REFERENCE/DOCKET NUMBER: C-37,075C
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (517) 636-8104
 : INFORMATION FOR SEQ ID NO: 11:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 795 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: join(144..192, 385..735)
 : FEATURE:
 : NAME/KEY: sig_peptide
 : LOCATION: join(144..192, 385..395)
 : FEATURE:
 : NAME/KEY: mat_peptide
 : LOCATION: 396..735
 : US-08-822-028-11

Query Match 0.5%; Score 16; DB 2; Length 795;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1061 tccactcaccacaca 1076
 |||
 Db 4 TCCACTCTCACACACA 19

Search completed: March 25, 2002, 19:18:53
 Job time: 21272 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 19:03:19 ; Search time 4399.74 Seconds

(without alignments)
11518.711 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcataagagcaaa.....ctttaactagtaactgct 3072

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3072	100.0	3133	9	AY032589	AY032589 Homo sapi
2	2919	95.0	3219	9	AY035391	AY035391 Homo sapi
3	2868	93.4	3396	9	AY027787	AY027787 Homo sapi
4	2868	93.4	3581	9	AF376061	AF376061 Homo sapi
5	1946	63.3	160583	2	AC010968	AC010968 Homo sapi
6	1844	60.0	138909	9	CNS01DS3	AL121653 BAC seque
7	1267	41.2	1355	9	IR2005417	AL121653 BAC seque
8	815	26.3	1395	9	AY027788	AY027788 Homo sapi
9	461	15.0	768	9	AY027789	AY027789 Homo sapi
10	292	9.5	162692	9	CNS01DS8	AL121658 BAC seque
11	292	9.5	185281	2	AC011232	AC011232 Homo sapi
12	269	8.8	578	9	AY027790	AY027790 Homo sapi
13	170	5.5	553	11	G55568	G55568 SHGC-100923
14	23	0.7	175152	2	AC074195	AC074195 Homo sapi
15	23	0.7	183556	2	AC019059	AC019059 Homo sapi
16	23	0.7	188459	2	AC090582	AC090582 Homo sapi
17	23	0.7	204487	2	AC090559	AC090559 Homo sapi
18	22	0.7	87834	9	AP002857	AP002857 Homo sapi
19	22	0.7	151088	9	AC020917	AC020917 Homo sapi
20	22	0.7	151366	33	AC007951	AC007951 Homo sapi
21	22	0.7	153733	2	AC011980	AC011980 Homo sapi
22	22	0.7	155531	2	AL592213	AL592213 Homo sapi
23	22	0.7	156230	2	AP002958	AP002958 Homo sapi
24	22	0.7	159946	2	AC079194	AC079194 Homo sapi
25	22	0.7	186413	2	AL161913	AL161913 Homo sapi
26	22	0.7	194533	2	AL592438	AL592438 Homo sapi
27	21	0.7	7595	4	BTCASK35	X14908 Bovine gene
28	21	0.7	37906	9	AC005199	AC005199 Homo sapi
29	21	0.7	45496	2	AC017948	AC017948 Drosophill
30	21	0.7	76526	2	AC027055	AC027055 Homo sapi
31	21	0.7	80346	8	ATM4E13	ATM4E13 Homo sapi
32	21	0.7	84499	8	ATT1205	AL022023 Arabidops
33	21	0.7	86710	9	ATFE3BE12	AL035522 Arabidops
34	21	0.7	130117	8	AC004907	AC004907 Homo sapi
35	21	0.7	157308	2	AC079351	AC079351 Homo sapi
36	21	0.7	183839	2	AC024590	AC024590 Homo sapi
37	21	0.7	184864	2	AC013553	AC013553 Homo sapi
38	21	0.7	194874	2	AC080090	AC080090 Homo sapi
39	21	0.7	195165	2	ATCHRIV82	AL161586 Arabidops
40	21	0.7	195217	2	AC025566	AC025566 Homo sapi
41	21	0.7	197859	8	ATCHRIV83	AL161587 Arabidops
42	21	0.7	198935	2	AC068573	AC068573 Homo sapi
43	21	0.7	206137	2	AL593857	AL593857 Mus muscu
44	21	0.7	211030	2	AL513468	AL513468 Mus muscu
45	21	0.7	221341	2	AC092892	AC092892 Homo sapi
46	21	0.7	227194	2	AC020727	AC020727 Homo sapi
47	21	0.7	303367	3	AE003538	AE003538 Drosophill
48	20	0.7	768	9	AY027789	AY027789 Homo sapi
49	20	0.7	865	2	AC078439	AC078439 Glardia i
50	20	0.7	1092	2	AC047412	AC047412 Glardia i
51	20	0.7	1198	14	REOSJNSB	M18390 Reovirus se
52	20	0.7	1688	8	SCYBR141C	Z36011 S.cerevisia
53	20	0.7	2833	8	SCYBR142W	Z36011 S.cerevisia
54	20	0.7	3709	3	AF116341	AF116341 Drosophill
55	20	0.7	12595	8	SCIRAL	X78937 S.cerevisia
56	20	0.7	36631	9	HUMCOL7A1X	L23982 Homo sapien
57	20	0.7	40937	3	CER01E6	Z68118 Caenorhabdi
58	20	0.7	45459	2	AC006103	AC006103 Homo sapi
59	20	0.7	83969	9	AC005210	AC005210 citb_179_
60	20	0.7	86155	9	AL159169	AL159169 Human DNA
61	20	0.7	86719	9	AP000885	AP000885 Homo sapi
62	20	0.7	88326	3	AC005923	AC005923 Homo sapi
63	20	0.7	88839	3	AC084447	AC084447 Caenorhab
64	20	0.7	91733	2	AL391278	AL391278 Homo sapi
65	20	0.7	93409	2	AC0733294	AC0733294 Mus muscu
66	20	0.7	102165	2	AC083819	AC083819 Mus muscu
67	20	0.7	106601	9	AL357352	AL357352 Human DNA
68	20	0.7	109810	9	AL139036	AL139036 Human DNA
69	20	0.7	114169	9	AC010902	AC010902 Homo sapi
70	20	0.7	120733	9	AC022124	AC022124 Homo sapi

QY 661 ctggatatacctggacacatcaggaaagacagacatcattcattgctcgtctgaagctgcg 720
|||||
Db 666 CTGGATATACCTGGACAAATCAGAGAGACGACATTTCATGCGCATGCTGCTGAAGTGGGG 755
QY 721 cagagggtctcttctcctctctgtatgctacaaatgaatcaggccccagaacttgcccagaa 780
|||||
Db 756 CAGAGGGGTTCTTTTCCCTTCTTGATGGCTACAAATGAATCAAGCCCCCAAGACTGGCCCAAA 815
QY 781 atcgaagccctcgataaagaacacacccgcttcaagaaactggtcatctgtaccacacac 840
|||||
Db 816 ATCGAAGCCCTGATTAAGAAAGAAACCCAGCTTCAAGAAATGATCATGTCACCACTAAC 875
QY 841 actgagtgctcctgaggacacatacggcaggtttggtccctcgactgctcgaagtggggata 900
|||||
Db 876 ACTGAGTGCCTGAGGCAATACGCGAGTTTGGTCCCTGACTGCTGAGAGTGGGGATATG 935
QY 901 acagaagaagcgcccaaggtctcactccagagaagtgtcgtatcaagaagagctgtctgaagg 960
|||||
Db 936 ACAGAAAGACGCGCCAGGCTCTCATCCGAGAAAGCTGATCAAGAGACTTGCTGAAGGC 995
QY 961 ttgttgctccaattcagaataaccaggtgctctgaggaaatccatgaagaacccctctctt 1020
|||||
Db 996 TTGTTGCTCCAAATTCAGAAATCCAGGTGCTTGAGGAATCTCATGAAGACCCCTCTCTT 1055
QY 1021 gttgttcactctgtgcaatccagatgggtgaagtgagtgctcactccacacacaaaca 1080
|||||
Db 1056 GTGGTCATCACTGTGCAATCCAGATGGGTGAAGTGGTTCACACTCCACACCAAAACA 1115
QY 1081 acgctgttccaatacctcctatgactcgttgaacagaaacaaacacacataaaggt 1140
|||||
Db 1116 ACGCTGTCCATACCTCTTATGATCTGTTGATACAGAAAAACAAACATATTAAGT 1175
QY 1141 gtgtgtcgaagtgactcattcctggagcctgagccactgtggaagcactgctcgtgaaggt 1200
|||||
Db 1176 GTGGCTGAAGTGAATCTTATTCAGAGCCTGAGCACTGTGAGACCTTACCTGTGAGGGT 1235
QY 1201 gtgtctccccaaggttgaattcgaactgagagatgctgcacgctgaagaagatctc 1260
|||||
Db 1236 GTGTTCTCCCAAGTTTGATTTGGAACCTGACAGATGCTCCAGGTGAATGAGGATCTC 1295
QY 1261 ctgtgtgaactggtgctcctctgtaaaatacagaactcaaaaggttcaagccaaagtataa 1320
|||||
Db 1296 CTGCTGACAACTGGGCTCTCTGTAATATACACTCAAAAGTTCAAGCCCAAGATATAA 1355
QY 1321 tctcttcaaggtcattccagaagtacacagcaggaagaaacacagcaggttattagcg 1380
|||||
Db 1356 TTCTTTTCACAGTCAATTCAGAGAGTACACAGCAGACGAATCAGAGATTATGACG 1415
QY 1381 tctcatgagccagagaggtgaccaaagggaatggttactctgcagaaatggttccatt 1440
|||||
Db 1416 TCTCATGAGCCAGAGAGGTGACCAAGGGAAATGTTACTTGCAAGAAATGTTCCATT 1475
QY 1441 tcggaattatcaccaactatagagcctgctccggtacacctgtgggtcatctgtgaa 1500
|||||
Db 1476 TCGACATTAATCAATCCACTTAAGACGCTGCTCCGGTACACCTGTGGGTCACTGTGAAA 1535
QY 1501 gccacgaaggtgttgaagaacacactcgacagcaggtgatacaacggtgtgctctcgtga 1560
|||||
Db 1536 GCCACACAGGCTGTTATGAAACACCTCGACAGATGATCAACAGGGGTGCTCTTGGA 1595
QY 1561 ctctccatcgccaagaggtccctctctgagagacaggaatcttgcagaagtgtgaaaaccc 1620
|||||
Db 1596 CTTTTCATCCGCCAAGAGGCTCTCTGAGACAGAAATCTTGCAAAAGTGTAAAAACCC 1655
QY 1621 actgaagcaagaattctgaaagcaataacatcaatccttctgttagagtggtgcacat 1680
|||||
Db 1656 ACTGAGCAAGAAATTCGAAAGCCATAAACATCAATCTTTGTAGAGTGTGCGCATCAT 1715
QY 1681 ttatatcaagaggttataccaataacagccctgagccaaagaaatttgaagcttcttcaa 1740
|||||
Db 1716 TTAATATCAAGAGATACATTCAAATCAAGCCCTGAGCCCAAGAAATTTGAAGCTTTCTTCAA 1775
QY 1741 ggtataagctatataatacaactcaggaacaatcccgatatacttacttacttcttga 1800

|||||
Db 1776 GGTAAACCTTATATATCACTACAGGAACAATCCCGCATTTACTTATTTGACTCTTTGAA 1835
QY 1801 catttgcccaattgtgcaagtctctggaactcattcaataactggaacttattggygaagct 1860
|||||
Db 1836 CATTTGCCCAATTTGGCAAGTGTCTGTGACATTTCAATTAACGTGACTTTTATGGGGAGCT 1895
QY 1861 atggtctatggaaaaggtctgcagaagaacaggttgaaatccacatgaagaagggccca 1920
|||||
Db 1896 ATGGCTTATGGAAGAAAGGCTGCGAAGACACAGGTGGAATCCATGGAAGAGGCCCA 1955
QY 1921 gaaacctacatccccaagaggtctgatacttctgtcttccaaactggaagcaggaattcag 1980
|||||
Db 1956 GAAACCTCATTTCCAGAGGCTGTATCTTGTCTTCAACTGGAACAGGAATTCAGG 2015
QY 1981 actctggaagtcaacactccggatctcagcaagtgtgaataagaagataacacatctg 2040
|||||
Db 2016 ACTGTGAGGTTCACACTCCGGGATTTGACCAAGTTGAATGAAGCAATATCATAATCTG 2075
QY 2041 gggaaaataatcagctctgcacaaagcctcaggtctgcaaatgaagaaggtgctggtg 2100
|||||
Db 2076 GGGAAATATTCAGCTCTGCCACAAAGCTCAGGCTGCAATTAAGAGATGCTGAGTGTG 2135
QY 2101 gctggaagcctcagtttgttctcagcaccctgtaagaacattatctctcatgtgtgaa 2160
|||||
Db 2136 GCTGGAACCTCAGTTTGTGCTCAGACACCTGTAAGAACATTTATTCATGATGTTGAA 2195
QY 2161 gccagctccctcaccacatagaagaatgagagacatcacatctgttaacaaactgaaaaac 2220
|||||
Db 2196 GCCAGTCCCTCACCATTGGAAGATGAGAGGCACATCATCTGTAACAAACCTGAAAAAC 2255
QY 2221 ttgagttatcaagctcacaagaatcaaggtctgcgggtggtctgacagaaagttggt 2280
|||||
Db 2256 TTGAGTATTCATGACCTTACAAATCAACAGCTGCGGGTGTCTGACGACAGCTTGAGT 2315
QY 2281 aacttgaagaaccttacaagaactcataatgataacataaagatgaatgaagaagatgct 2340
|||||
Db 2316 AACCTGAAGAACCTTTACAAACCTCATATGATGAATGAATGAATGAAGAAAGATGCT 2375
QY 2341 ataaactagctgaagcctgtaaaaaacctgaagaagatgtgttatttcaatttgaaccac 2400
|||||
Db 2376 ATAAACCTAGCTGAAGGCTGTAAGAAACCTGAAAGAGATGTTTATTTCAATTTGACCCAC 2435
QY 2401 ttgtctgaacttgaagaaggaatgattacataatgaagttctcgtgaaggaacccgt 2460
|||||
Db 2436 TTGTCTGACATTTGAGAGGATGATTAATGATGATGATGATGATGATGATGATGATGAT 2495
QY 2461 gaccttgaagaatcattatgctcctcgtcgtctgtctgtcgaatgcaatgcaatgcaatccta 2520
|||||
Db 2496 GACCTTGAAGAAATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2555
QY 2521 gctcaagaacttcaacaattgtgtcaaacatgagacatcttattatcaagaataactcgt 2580
|||||
Db 2556 GCTCAGAAATCTTCAAAATTTGTCAAACTGAGCAATTTGATTTATTCAGAAATTTACCTG 2615
QY 2581 gaaaaagaatgaagaagccttcaatgaacttgatgaagaagatgaagaaggtctgaagacag 2640
|||||
Db 2616 GAAAAAGTGAANAATGAAGCTCTTCAATGAATGATGACAGAGATGAAGCTCTGAAACAG 2675
QY 2641 ctcaaccgaactgctgcccctggtggtctgtgacgtgcaaggaagcctcgaagcagcctgtg 2700
|||||
Db 2676 CTCACCGCATGATGCTGCTGGGGGCTGTGACGAGCAAGGACACCTGAGCAGCTGTGG 2735
QY 2701 aaacaatttgaagaaggtcccaaacactgctcaagcttgggttgaanaactggaagacaca 2760
|||||
Db 2736 AAACATTTGGAGGAGTCCCAACACTGCTCAAGCTTGGGTGAAAGAACTGAGAGACTACA 2795
QY 2761 gatacagagattgaagaatttgaaggtgcaatttttgaagaagaacccctgaaaaacttcag 2820
|||||
Db 2796 GATACAGAGATTAATAATTTTATGATGATTTTGTGAAAGAACCTCTGAAAGAACTTCAG 2855
QY 2821 cagttgaatttgcgggaataatcgtgtgagcagtgatgaatgctctgctcatatgtgtga 2880
|||||

Db 2856 CAGTTCGATTGGCGGGAATCGTGTGAGCAGTGATGGATGCTTGCCTTCATGGGTGA 2915
 QY 2881 ttgagaatcctaagaactagtgcttttcgacttagtaactaagaattcactcgat 2940
 Db 2916 TTGAGATCTTAGCAATAGTGTGCTTTGACTTAGTACTAAGAAATTTCTACCGAT 2975
 QY 2941 ccagcattagtcagaaactaagccaagtgcttccaaagtaactcttcgcaagaagct 3000
 Db 2976 CCAGCATTTAGTCAGAAAACCTTACCAAGTGTATCCAAAGTAACCTTTTCGCAAGAAGCT 3035
 QY 3001 aggcctgttggtggtgcaacttagatgatgatcagtgcttcttaacaggtgcttttaa 3060
 Db 3036 AGCCTGTGTGGTGCGCAATTTGATGATGATGATCTCAGTGTATTTACAGGTGCTTTTAA 3095
 QY 3061 ctgaactagct 3072
 Db 3096 CTGATGACTGCT 3107
 RESULT 2
 LOCUS AY035391 3219 bp mRNA PRI 24-JUL-2001
 DEFINITION Homo sapiens ICE-protease activating factor mRNA, complete cds.
 ACCESSION AY035391
 VERSION AY035391.1 GI:14334214
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3219)
 AUTORS Poyet,J.L., Stinivasula,S.M., Thnani,M., Razmara,M.,
 Fernandes-Alnemri,T. and Alnemri,E.S.
 TITLE Identification of Ipaf, a human caspase-1-activating protein
 related to Apaf-1
 JOURNAL J. Biol. Chem. 276 (30), 28309-28313 (2001)
 MEDLINE 21359454
 PUBMED 11390368
 REFERENCE 2 (bases 1 to 3219)
 AUTORS Poyet,J.-L., Stinivasula,S.M., Fernandes-Alnemri,T. and
 Alnemri,E.S.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
 Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
 USA
 FEATURES
 source Location/Qualifiers
 1..3219
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="p21-22"
 145..3219
 /note="IPAF: CED4/Apaf-1 family member: caspase-associated
 recruitment domain containing protein"
 /codon_start=1
 /product="ICE-protease activating factor"
 /protein_id="AAK59843.1"
 /db_xref="GI:14334215"
 /translation="MNFIKDNRSLALFORMGMVYIKITDLPFVNVNLRNEVINICE
 KYEODARGLIHMLIKKSGESCNPLFKSKENNYPLFOLNLSALPHONSPGDIIDLA
 ODLDKDIHTPSELNTPYLEDIDITFNLSSTPEYPLMKDKDHHNEVDLTGLGLQA
 LOSPCITIGESGKSTLQRIAMLWGSCKALTKFKVFPFLRUSRAAGGLFETLCD
 OLDDIPGTIRKQTFAMLLKRLQRYVFLDGVNEFPQNCPELEALIKENHREKNVI
 VTTTTECLRHIFQFALTAEGVMDTDSQALIREVILKEGLLQIOKSRCLNL
 MTPPLEFVITICAIOMGESEFHSHTOTTFEFTFDLLIQNKHKHKAASDFRSLDH
 CGDLAEGVFSKHFDELODVSVNEDVLTGLTGLCKYTAORPKPYKRFPHKSPQYT
 AGRLSSLLTSHPREVTGKNTLOKMYISDITSTYSLLXRTCCSSVEATRAYVKH
 LAAVVOHGLGLSLIAKRLPMKQESLQSVKNTTDEILAIAINNSVVEGILHYOST
 SKSALSGEFAFQGSILYINGNIPDYLFDEFENPCASALDFIKIDFYGGAWSW
 EKAADETGGIIMEEAPETLYIPSKAVSLFPMKQEFPLEVTLADFELKDKODIRYLG
 ITESATSLRQIKRCAGVAGSLPLVTKKNITSLTVEERHITSTVNTLT
 LSHLDONORLPQGLTDSIGNIKNTLKLMDINKMEEDAPILAEIKLKKMCPHLL
 THLSDIGEMDIYKLSLSEPCDLEIQLVSCCLSNANAYKILANQNLHNLYKSLIDLS

ENVLEKDGNEALHELIDRNNVLEQLTALMLPNCQVQSSLSLKHLEBPOLYKLG
 KMWRLDEIRILGAFPGKRNPLKFNQOQLMIAQNRVSDGMFAFMGVENLQOLFEDF
 SYKELEPDPALVRKLSOVLSPKLPFQEARLVGMQFDDDLISVITGAFRLVTA"
 BASE COUNT 934 a 696 c 748 g 841 t
 ORIGIN
 Query Match 95.0%; Score 2919; DB 9; Length 3219;
 Best local Similarity 99.9%; Pred. No. 0;
 Matches 3069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 atgaattcctaagaacaaatagccgaagcccttattccaagaatggaatgactgtata 60
 Db 145 ATGAATTTCAATTAAGACAAATATACCCAGCCCTTATTCMAAAGATGGCAATGACTGTATA 204
 QY 61 aagcaatcacagatgacccattgtatgaaatgcttgaatcgagcaagaatgaacatc 120
 Db 205 AACCAATCACAGATGACCTATTTGTATGGAATGTTCTGATGCCAGAGAAGTAACATC 264
 QY 121 attgtctgagaaggtgagcagatgctgtagaggatcatatcacatgatttga 180
 Db 265 ATTTGCTGCAGAGAAGGTGAGAGATGCTGTAGAGGATCATTCATGATTTTGAAA 324
 QY 181 aaggttcaagatccgtgaaccccttcttaaatcccttaaggatggtgaactctcta 240
 Db 325 AAGGTTCAAGATCCGTAACTCTTTTAAATCCCTTAAAGAGATGGAACTATCCCTTA 384
 QY 241 ttccgaacttgatgacaaagctcttttcacagacatcacagaaggaactgagcagat 300
 Db 385 TTTCAGACTTGAATGACCAAAAGCTTTTTCATCAGACATCAAGAGAGACTTGGACGAT 444
 QY 301 tggctcagatllaaaggactgtacacacacacatcttctgaacttataccctt 360
 Db 445 TTGGCTCAGATTTTAAAGCACTGTACCATACCCCATCTTTCTGAACCTTTATCCCTT 504
 QY 361 ggtgaagatatgaactattttaaacttgaagaagaccccttccagaacctgtccgtg 420
 Db 505 GGTGAAGATATTAACATTAATTTTAACTTGAAGAACACCTTCAACAACCTGCTGTGG 564
 QY 421 aggaagacacaacacacacacacgcgtggagcagctgaccccttgatgctccctgcagct 480
 Db 565 AGGAAGACCAACACCATCACCGCGTGAGACACTGACCTGATAGGCTCCGCGACAGGCT 624
 QY 481 ctccaagccccctgcacatcattgaagggaactctgcgaaggcaagctccactgtgcag 540
 Db 625 CTTCAGAGCCCTGCAATCATTTGAAGGGGAATCTGCAAGAGCAAGTCCACTGCTGCAG 684
 QY 541 cgcattgccatgctctggggctccggaaggtgcaaggctctgcgaagtccaatgcgtc 600
 Db 685 CGAATTCCTGCTGTGGGCTCCGGAAGTSCAAGGCTCTTGACCAAGTTCAATTCGTC 744
 QY 601 ttctccctcgctcagcaagggcccaagggtggaacttttgaacccctctgtacaactc 660
 Db 745 TTCTTCCTCGCTCAGCAGAGGCCCAAGGTGACCTTTTCAAAACCTCTGTGATCAACTC 804
 QY 661 ctgatatactcggcaacatcaaggaagcagacatcaatgagcatgctgtcgaagctgcg 720
 Db 805 CTGATATACCTTGCAATCAGGAAGACATTTATGCGCATGCTGTGAAGCTGCGG 864
 QY 721 caaagggtcttcttcttctgtgtactaagaatlaaagcccaagactgcccagaa 780
 Db 865 CAAAGGCTTTCTTCTCTTGTATGCTACAAAGAAATTCAGGCCCAAGACTGCCAGAA 924
 QY 781 atggaagccctgtaaaagaaacacccgcttcaagaacatgtgcatcgtaacatacc 840
 Db 925 ATGGAAGCCCTGATTAAGGAACACACCGCTTCAAGAACATGCTCATCGCACCATAC 984
 QY 841 actgaatgctcgaagcacaatacagcaagttgtgtccctgaactcgtcgaagtggggatag 900
 Db 985 ACTGAGTGCCTGAGGACATACAGGAGTTGTGGCTGTGACTGTGAGGAGGGGATATG 1044
 QY 901 aacgaagacagcccaaggtctcattccgagaaggtgtctgaataagagcttgtgaagc 960

Db	1045	ACAGAAAGACAGGCCAGGCTCTCATCCGAGAGTGTGATCAAGAGCTTTCGAAAGGC	1104
Oy	961	ttgtgtcccaaatccagaataatccagatgctgtgaagaatctcaataagaacccctctctt	1020
Db	1105	TTGTGTCTCCAAATTCAGAAATCCAGGTGCTTTGAGGAATCTCATAGACCCCTCTCTTT	1164
Oy	1021	gtgtcatcacttgtgcataatccagaatgggtgaagtgagttccactctcacacacaa	1080
Db	1165	GTGTGATACACTGTGTGCATCCAGATGGGTGAAGTGAAGTTCACCTTCACACACAAACA	1224
Oy	1081	acgcgttccataactctctatgactgtgtgatacagaataaacaacacacataaagtc	1140
Db	1225	ACGCTGATCCATACCTTGTATGATCTGTGATACAGAAAAACAAACAAACATTAAGGT	1284
Oy	1141	gtgagtgaagaatgacttcatctcgagacactgtgagacactgtgagacactgtctgaaggt	1200
Db	1285	GTGGCTGCAGAGTACTTATTCGGAGCCTGGACACTGTGGAGACTGTGAGGTGAGGCT	1344
Oy	1201	gtgtctcccaaatcttgatcttcgaactgcagagatgtctccagcgtgaatgaagatgtc	1260
Db	1345	GTGTCTCCCAAGTTTGATTTGCACTGCAGATGTGTCCAGCGTGAATGAGATGTCTC	1404
Oy	1261	ctgtgtacaactggcctcctctgtaataataacagctcaaaagttcaagccaaagtataa	1320
Db	1405	CTGCTGACAACTGGGCTCCTGTGTAATATACAGCTCAAAAGGTTCAAGCCAAAGTATAA	1464
Oy	1321	ttcttccaagaatctccaagaatgatacagcagagagaagaactgaagacttataatgagc	1380
Db	1465	TTCTTTCAAGTCAATTCAGAGATACACAGCAGACAGCAAGCTCAGAGATTATTTAGCG	1524
Oy	1381	ttctcaatgacgaagaatgatacagaagaatgatacagaagaatgatacagaagaatgatac	1440
Db	1525	TCTCATGAGCCAGAGAGAGTGAACAGAGGGAATGTTACTTTCAGAAATATGTTTCCATT	1584
Oy	1441	tcgagacatcaatcaactataagacagcctgcgcgtacacactgtggatcgtgtgaa	1500
Db	1585	TGCGCATTTACATCCACTTAATAGAGCCTGCTCCGCTACACCTGGGCTCATCTCTGAAA	1644
Oy	1501	gccacgaaggctgttataagaacactgcagcagatgatacaagaagctgtcctctcga	1560
Db	1645	GCCACAGGGCTGTTATAGAGCACTCCGAGAGTGTATCAACAGCGCTGCTTCGGA	1704
Oy	1561	cttctccatgcgaagaagcctctctgcagagacagaatcttgcgaagtgtgaaataaac	1620
Db	1705	CTTTCATTCGCCAAGAGGCTCTCTGAGACAGGAATCTTTGCAAAAGTGAATAAACCC	1764
Oy	1621	actgagcagaagaatctgtgaagacataaacaatccatctctgttagagttfgygacat	1680
Db	1765	ACTGAGCAAGAAATTTCTGAAAGCCATAAACATCAATTCCTTGTAGAGTGTGGCATTCAT	1824
Oy	1681	ttatacagaagaatgatacacaatcagcctgcagcagaagaatgtgaagcttcttcaa	1740
Db	1825	TTATATCAAGAGAGTACTTCACAAATCACCCCTGAGCCAGAAATTTGAAAGCTTCTTTAA	1884
Oy	1741	ggtaaagcttatatatcaactcagagagacalcccgaattacttaattgacttcttga	1800
Db	1885	GGTAAAGCTTATATATCACTACAGAGAACATPCCCGATTAATTGACCTCTTTGAA	1944
Oy	1801	catctgcccgaattgtgaagtgctctgtgaactcatataactgagacttatafgggagct	1860
Db	1945	CATTGGCCCAATTTGCAAGAGCCCTGAGCTTCATTAACTGAGACTTTTATGGGGGAGCT	2004
Oy	1861	atggtctcataggaagaagctgcgaagaacagaatgtgaatccacatagaagaagcccca	1920
Db	2005	ATGGCTTATATGGAAAGGCTGCAAGAGACACAGGTGGAATCCACATGGAAGAGGCCCA	2064
Oy	1921	gaacatcaatctccagaagggctgtatcttgtctctcaactggaagcagaagaatcag	1980
Db	2065	GAAACCTACATTTCCAGAGGGCTGTATCTTGTCTTCAACTGGAAGCAGGAATTTACAG	2124
Oy	1981	actctggagtgcaaacactccggatcttcagcaagtgtgaataagaatatatacatatctg	2040
Db	2125	ACTGTGAGGATCACACTCCGGGATTTTCAGCAAGTTGATTAAGCAAGATATCAGATATCTG	2184

Oy	2041	gggaaatatatcagctctgcacaaagcctcaagcgtcgaataataaagatgtgtcgtgtg	2100
Db	2185	GGGAAATATTCAGCTCTGCCACAGCCTCAAGCTTCAGAAATTAAGATGTGCTGTGTG	2244
Oy	2101	gtctgaagcctcagtttgcctccagaccctgttaagaacattatctctcaatgtgtgaa	2160
Db	2245	GCTGGAACCTCAGTTTGTGCTCCTCAGCACCTGTGAAGCAATTTATTTCTCATGTGTGAA	2304
Oy	2161	gccagttcccctcaacatagaagaatgagagacatcacatctgttaacaacctgaaacc	2220
Db	2305	GCCACTCCCTCACCATTTGAAGATCAGAGCACATCAATCATCTGTAAACAACCTGAAAAAC	2364
Oy	2221	ttgagttatctagacctcaagaatcaagcgtcggcgggtgtgtctgacgtgaagcttgggt	2280
Db	2365	TTGAGTATTTCAATGACTACAGAAATCAAGGCTGCCGGGTGTGTGACTGACAGCTTGSGT	2424
Oy	2281	aactggaagaacctatacaagaactcaataatgataacataaagaatgaatgaagaatgct	2340
Db	2425	AACTTGAAGAACCTTACAAAGCTCATATGATGATTAACATTAAGATGAATGAAGAAATGCT	2484
Oy	2341	ataaaactagctgaagcctgtgaaaaacctgaagaagatgtgttaattcaattgaaccac	2400
Db	2485	ATTAACACTAGCTGAAGCCTGAATAAACCTGAAGAAGATGTATTATTTCAATTTGACCCAC	2544
Oy	2401	ttgtctgacattggaagggatgattacataatgataagctctcgtcaagtgaaacctgt	2460
Db	2545	TTGTCTGACATTTGAGAGGAGATGGATGATACATATGACAGTCTGTCTGTAAGTAACCCGT	2604
Oy	2461	gaccttgaagaatctcaatgactcctcctgcgtgtgtctgtcaaatgcaatgaataactca	2520
Db	2605	GACCTTGAAGAAATTCATTTAGTCTCCTGCTGCTGTGTGTGCAATGCAATGCAATGCTTA	2664
Oy	2521	gtctgaactcttcaaatctgtgtcaaacctgagcatctctgtaattatcaagaataactctg	2580
Db	2665	GCTGAGATCTTCAACAATTTGTGTCAAACTGAGCAATCTTGTGATTTATCGAAAAATTTACTCG	2724
Oy	2581	gaaaagaatggaagaatggaagcctctcatatgaactgtatcagcagaagaatgaagctctcga	2640
Db	2725	GAAAAAGATGGAATGAAGCTCTTCATATGACTGATGACAGAGATGAAAGTCTGAAACAG	2784
Oy	2641	ctcacccgacatgactgtgcccctggtggtgtgtgacgtgtgcaagacagcctgaagacactgtg	2700
Db	2785	CTCACCGCAGATGATGCTCCTGGGGCTGTGACGTCGAAGCAGCGCTGACGACGCTGTG	2844
Oy	2701	aaacatttggaggggtgtcccaactgtcaagcttgggtgtgtgaaaacctggaagactaca	2760
Db	2845	AAACATTTGGAGAGGTCCACAACTCGTCAAGCTTGGGTTGAAAAACTGAGAGCTCACA	2904
Oy	2761	gatacagaatataaattttagtgcattcttctggaagaagacccctcgaaaaacttcag	2820
Db	2905	GATACAGAGATTAGAAATTTTAGGTGTCATTTTGTGAAGAAACCTCTGAAAAACTTCACAG	2964
Oy	2821	caatgtgaatttggcgggaataatcgtgtgacagtgatgtgaatgtgtcctcctcaatgtgtca	2880
Db	2965	CAGTTGAATTTGGGGGAATATGTCGTGAGCAGTGTGATGATGGCTTGCCTTCATGGGTGA	3024
Oy	2881	tttgaagaactcttaagcaatgagtggttttctgacttagtactaaagaatcttcaactgat	2940
Db	3025	TTTGAAGATCTTAAAGCAATTTAGTGTATTTTGTGACTTTAATACTTAAGAAATTTCTACCTGAT	3084
Oy	2941	ccagcatagtcagaaaacttaagccaagtgtataccaagttaacttctctcgaagaagct	3000
Db	3085	CCAGCATTTAGTCAGAAAACCTTACCAAGCTGTTATTCAGATTAATTTCTCAGAAAGACT	3144
Oy	3001	aggtctgttgggtggcaatttgatgatgatcctcagatgtgtatatacaggtgactttaaa	3060
Db	3145	AGGCTGTGTGGTGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3204
Oy	3061	ctagttaactgct 3072	
Db	3205	CTAGTAACCTGCT 3216	

LOCUS	3	20-JUL-2001
AY027787		
DEFINITION	3396 bp	PRI
ACCESSION	Homo sapiens CLAN1 (CLAN1) mRNA, complete cds.	
VERSION	AY027787	
KEYWORDS	GI:14324112	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	1 (bases 1 to 3396)	
TITLE	Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.	
JOURNAL	Clan, a novel human ced-4-like gene	
MEDLINE	Genomics. 75 (1-3), 77-83 (2001)	
PUBMED	21365712	
REFERENCE	11472070	
AUTHORS	2 (bases 1 to 3396)	
TITLE	Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.	
JOURNAL	Direct Submission	
	Submitted (21-FEB-2001) Program on Apoptosis and Cell Death	

FEATURES	Location/Qualifiers
source	1. .3396

gene

CDS

BASE COUNT	992 a	737 c	793 g	874 t
ORIGIN				

Query Match	93.4%	Score 2868	DB 9	Length 3396
Best Local Similarity	99.9%	Pred. No. 0		
Matches 3068; Conservative	0	Mismatches	4	Indels 0
				Gaps 0

Qy	1	atgatattccataaagaganaaatagccgagacctttccaagaatagggaatgactgtgtata	60
Qy	277	atgattttcattaaagcaaatagccgagaccttttttcaaaagaaagggaatgactgttata	336
Db	61	aagcaatcacagatgacctatttgtatggaatggtttctgaaatcgcgagaagataaacatc	120
Qy	337	aagcaaatcacagatgaccttttttgtatggaatggtttctgaaatcgcgagaagataaacatc	336
Db	337	aagcaaatcacagatgaccttttttgtatggaatggtttctgaaatcgcgagaagataaacatc	336

[illegible]

|||||
Db 1477 GTGTTCTCCCAAGATTGATTTGCAACTGCAAGATGTGTCCAGCGTGAATGAGATGTC 1536
Qy 1261 ctgtcgaaacttgggtctctctgttaataatacagctcaaaagttccaagcaagtataa 1320
Db 1537 CTGTGAGAACCTGGGCTCCTCTGTAAATATACAGCTCAAGGTTCCAAAGATATATAA 1596
Qy 1321 tctcttcaagtcattccagagatlaacacagcagaagaactccaagttattttagcg 1380
Db 1597 TTTCTTTCAAGGATTTCCAGAGATACACAGCAGACAGCAAGCTCAGCATTTATGACG 1656
Qy 1381 tctcatgagccaagagagtgaccagaagggaatgttacttgcagaaaatgttccatt 1440
Db 1657 TCTCATGAGCCAGAGAGAGGTACCAAGGGGAATGTTACTTTGCAAGAAATGGTTTCCATT 1716
Qy 1441 tcggaacattacaactcaacttataagacgctgtccgtacacccgttgagtcattgtgaa 1500
Db 1717 TCGGACATTTACATCCACTTAATAGCAGCCTGCTCCGGTACACCTGTGCTCATCTGTGAA 1776
Qy 1501 gccaccaagggtgttattgaagcacctgcagcagctgtatlaacaacggtgtgctctcgga 1560
Db 1777 GCCACACAGGCTGTATTATGAAACACCTCGACAGATGTAACACAGGCTGCTTCTCGGA 1836
Qy 1561 ctctccatcgccaagagagcctctcttgagacaggaatcttgcagaagtgtgaaacaccc 1620
Db 1837 CTTTTCATCCGCCAAGAGGCTCTCTGGAGACAGAACTTTGCAAAAGTGTAAACACAC 1896
Qy 1621 actgagcaagaatcttgcagaacataacatcattccttctgtagagtgtgcacat 1680
Db 1897 ACTGAGCAAGAAATTTCTGAAAGCCATTAACATCAATTCCTTTGAGAGTGTGGCATTCAT 1956
Qy 1681 ttatatacagaagagtlacatccaatcaagccttgagccaagaatttgaagccttcttcaa 1740
Db 1957 TTATATCAAGAGATACATCCAAATACGCTTGAGCCAGAAATTTGAAGCTTTCTTCAA 2016
Qy 1741 ggtlaaagctatatalatcaactcagaggaacatcccgatlaactatlaacttgaacttctgaa 1800
Db 2017 GGTAAAGCTTATATATCACTCAAGGAAACATCCCGATTAATTTGACTTTCTTGA 2076
Qy 1801 catttgcacatttgcagaatgtcctgtgacttcaacttgaagccttcttgaagagct 1860
Db 2077 CATTTGCCCAATTTGTGCAAGTGCCGTGACATTTCAATTAACGTGACTTTATGGGGGACT 2136
Qy 1861 atggtctcatgaggaagagctgcagagaacacaggttgaatccacatgagaagagcccca 1920
Db 2137 ATGGCTTCATGGAAGAGGCTGCAAGACACAGGTGAATCCCATGGAAGAGGCCCA 2196
Qy 1921 gaaacttcaacttccacagcagagctgtatcttcttcaacttgaagcaggaattcag 1980
Db 2197 GAAACCTACATTTCCAGAGGCGTGTATCTTGTCTTCAACTGGAAACAGGAATTCAGG 2256
Qy 1961 actctgagagtcacacactccgggatttcaagaagttgaataagcaagatalcatatctg 2040
Db 2257 ACTCTGGAGGTACACTCCGGGATTTTCAGCAAGTTGAATAGCAAGATATCATATCTG 2316
Qy 2041 gggaataatcagctctgcagacaagcctcaggtctgacaataaagaagatgagctggtg 2100
Db 2317 GGGAAATATTCAGCTCTGCCACAAAGCCTCAGGCTGCAAAATGAAGATGTGCTGTG 2376
Qy 2101 gctggaagcctcagtttgttctcagcagcctgttaagaacatttctctcatatgtgaa 2160
Db 2377 GCTGGAAGCCTCAGTTGGTCTCTCAGACCTGTGAAGAACTTTATCTCTATGCTGAA 2436
Qy 2161 gccagctccctcaccatagaagaatgagaagcacaatcactgttaacaacctgaaaac 2220
Db 2437 GCCAGTCCCTCACCATTAAGAGATGAGGCAATCAATCTGTAAACAAACCTGAAACCC 2496
Qy 2221 ttgagttatcagcctcaccagaatcaaacggtgcgggtgtgtcgaacgagagcttgggt 2280
Db 2497 TTGAGTATTCATGACCTTACAAATCAACGCTGCCGGGTGTGTGACTGACAGCTTGAGT 2556
Qy 2281 aacttgaagaaccttacaagctcaaatgtatatacaataaagatgaatgaagaatgct 2340
|||||

Db 2557 AACTTGAAAGCACTTACAAAGCTCATATGATATACATTAAGATGAATGAAGATGCT 2616
Qy 2341 ataaactagcttgaagccttgaaaaaacttgaagaagatgttatttcaatttgaaccac 2400
Db 2617 ATAAAGTAAGCTGAAGGCTTAAAAAAGAGAGATGTTATTTCAATTTGACCCAC 2676
Qy 2401 ttgtctgacatttgaagaagaaatgatataatgaacttctgtcaagtgaaacccgt 2460
Db 2677 TTGCTGACATTTGAGAGGGAATGATTAATAGTCAAGTCTTGTCAAGTGAACCTGCT 2736
Qy 2461 gaccttgaagaatccaattagctctcctgcgtctgtctgtcgaanaatgcagtgaaatccca 2520
Db 2737 GACCTGAAGAAATTCATTAATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2796
Qy 2521 gctcaagaatcttcaacaatttgttcaaacatgagcattcttgaattcaagaatlaactg 2580
Db 2797 GCTCAGAACTTTCACAAATTTGTCAAAGTGAAGCATTTCTTGAATTAAGAAATTTACCTG 2856
Qy 2581 gaaaaagatggaatgaagccttcaatgaacttgaacttgaagatgaagtgtagaagcag 2640
Db 2857 GAAAAAGATGGAATGAAGCTCTTTCATGATGATCCAGCAGATGAAGCTCTGAAACAG 2916
Qy 2641 ctcaacgactgactgtgccccttgaggctgtgacgttgcagaagcagccttgaagcctgtg 2700
Db 2917 CTCACCGCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2976
Qy 2701 aaacaattggaagaggtcccaacaactgtcaagcttgggttgaanaacttgaagactca 2760
Db 2977 AAACATTTGGAGGAGGTCCCAACCTGCTCAAGCTGTGGGTGAAGAAACTGAGACTCACA 3036
Qy 2761 gatacagaatgaagaatttgaagtgatcttcttgaagaagccttgaanaacttccag 2820
Db 3037 GATACAGAGATTAATTTTATGATGATTTTGGAAAGAACCTCTGAAAGAACTTCCAG 3096
Qy 2821 cagttgaatttgcgaggaacatgctgtgagcagtgatgagtgatgcttgccttcaatggtgta 2880
Db 3097 CAGTTGAATTTGGCGGGAATCGTGTGAGCAGTGTGATGATGATGATGATGATGATGATGAT 3156
Qy 2881 ttgagaatcttaagcaattggttcttcttgaacttgaacttgaagaatcttcaactg 2940
Db 3157 TTTGAGAACTTAAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3216
Qy 2941 ccagcatgactcagaagaacttgaagcagtgatlaaccaagttcaacttctgcagaagct 3000
Db 3217 CCAGCATTAAGTCAAGAACTTAAGCAAGTGTATTCAGAACTTTCTTCTCAAGAACT 3276
Qy 3001 agccttgttgggtgcaatttgaatgaatgaatcagtgatlaacttcaagtgctttaa 3060
Db 3277 AGGCTTGTGGTGGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3336
Qy 3061 ctgataactgct 3072
Db 3337 CTAGTAATCTGCT 3348

RESULT 4
AF376061
LOCUS AF376061 3581 bp mRNA PRI 15-MAY-2001
DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
ACCESSION AF376061
VERSION AF376061.1 GI:14040074
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3581)
Gingras, M., Qiu, J. and Margolin, J.F.
DIFFERENTIAL EXPRESSION OF THE CASPASE RECRUITMENT DOMAIN PROTEIN
12 (CARD12) DURING MONOCYTIC DIFFERENTIATION
JOURNAL
Unpublished
REFERENCE 2 (bases 1 to 3581)

AUTHORS . Gingras, M., Qiu, J. and Margolin, J. F.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center,
Baylor College of Medicine, 6621 Fannin St. MC3-3320, Houston, TX
77030, USA

FEATURES

Location/Qualifiers

1..3581

source

CDS

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p21-22"
/cell_line="U937"
/cell_type="peripheral blood-derived CD14 mature monocyte"
430..3564
/note="CARD12"
/product="caspase recruitment domain protein 12"
/protein_id="AAK5343.1"
/db_xref="GI:14040075"
/db_xref="GI:14040075"

misc_feature

502..741
/note="Region: caspase recruitment domain"

BASE COUNT 1033 a 781 c 843 g 924 t

ORIGIN

Query Match 93.4%; Score 2868; DB 9; Length 3581;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3068; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgaattcataagaagcaatagccagccctattcaagaatgggaatgacgtatata 60
DB 490 ATGAATTTCTAAGGACATAGCCGAGCCCTATTCAAGAAATGGGAATGACGTATATA 549
QY aagcaaatcacagatgacatcttctgtaagtctgtaacgcgagagaagtaaacatc 120
DB 550 AAGCAATACACAGATGACGATTTGTATGGAATGTTCTGATCGCGAAGAAATTAACATC 609
QY 121 attgctcgagagaagtgagagcagatgctgctagaaggaatcattcaatgatttgaaa 180
DB 610 ATTTGCGCGAGAGGAGGAGGAGATGCTGTAGAGGAGATCATTTCAATGATTTTGAAA 669
QY 181 aagagttcaagatccctgttaacctcttctaactccttaagagatggaacttcctcta 240
DB 670 AAGGTTTCAGAGTCTGTAACTCTTTCTTAATCCCTTAAGGAGTGGAATCATCCCTTA 729
QY 241 ttccagacttgaaatgagcaaaagtcttttcatcagacatcagaagagacttgagacat 300
DB 730 TTTCAGGACTTGAAATGACAAAGCTTTTTCATCAGACATCAGAGAGACTTGAGAGAT 789
QY 301 ttgagtcaggaatttaaagagactgtacacatcccatcttctgaactttatccctc 360
DB 790 TTGGCTCAGAGATTTAAAGGACTGTGTACATACCCCATCTTTTGAGACTTTATCCCTT 849
QY 361 ggtgaagatattgacattattttaacttgaagaagccttcaagagactgtccctgtg 420
|||||

DB 850 GGTGAAGATATTGACATTATTTTTAACTGTAAGAAAGCACTTCACAGAACCATGCTGTGG 909
QY 421 aggaagaccacaacacatcacccgctgtagaagctgaacctgaatgacctctgacagct 480
DB 910 AGGAGGACCAACACATCATCCCGCTGTGAGACACTGACCTGTAATGGCCCTCCGACGGCT 969
QY 481 ctccaagacccttgatcatcattgaagggaatctgtgcaaaagcaagtcacactgtgacg 540
DB 970 CTTACAGGCCCTGACATTCATTTGAAGGGGAATTCGGCAAAAGCAAGTCCACTGCTGCAG 1029
QY 541 cagattgccaatgctctggggctccggaagaatggaagctctgacagatcgaatcgtc 600
DB 1030 CGAATGCAATGCTCTG6GGCTCCGGAAGTGAAGGCTTGACCAAGTTCAATTCGTC 1089
QY 601 ttctccctccgtctcagcaagggccaggggtgagacttttgaacctctgtgatactc 660
DB 1090 TTCTCTCTCGTCTCAGCAGGGCCAGGGTGGAGCTTTTGAACCCCTGTGATCAACT 1149
QY 661 ctgatatatccttgacaaatcaaggaagcaagacatcaatgacatgctgtgagctgag 720
DB 1150 CTGGAATACCTGACCAATCAGGAAGCAGACATTCATGCGCATGCTGCTGAAGCTGCGG 1209
QY 721 caagaggtctcttctctctgtagtctataagaaatcaagccccaagactgccaaga 780
DB 1210 CAGAGGCTTCTTCTCTCTTGTGATGGCTCAATGAATTCAGGCCCAAGACTCCAGAA 1269
QY 781 atcgaagcccttgataaaggaagaacccagctctcaagaacatggtcgtcacactacc 840
DB 1270 ATCGAAGCCCTGATTAAGGAAGAAACCAACCGCTTCAAGAACATGTCATGTCACCATCAC 1329
QY 841 actgaatgcttgaggaacatacggcagcttggtgacctgacgtcgaagtggggatag 900
DB 1330 ACTGAGTGGCTTGAGGACACATACGCGAGTTGTGGTCCCTGACTGAGGGGGATATG 1389
QY 901 acggaagacagccagagctctcaccggaaggtctgataaggaagcttgtagaagc 960
DB 1390 ACGAAGACACGCCCCAGAGGCTCTCATCCGAGAGTGTCTATCAAGAGCTTGTGTAAGGC 1449
QY 961 ttgtgctccaatccagaatccaggtgcttgtaggaatcattgaagaccctctctt 1020
DB 1450 TTGTGTCTCAAAATTCAGAAATCCAGGTGCTTAGGAATTCATGAAGACCCCTGCTTT 1509
QY 1021 gtgtgatacactgtgcaatccagatgagtgaaagtgaatgccatctcacacaaaca 1080
DB 1510 GTGCTCATCTCTGTGCAATCCAGATGGTGGAAGTGAGTTCACACTCTCACACAAACA 1569
QY 1081 accgtgtccatccctctcgtatgctgtgatacagaagaaacaaacaaacaaataaggt 1140
DB 1570 ACGCTGTTCATACCTTTCTATGATCTGTTGATACAGAAAACAAACATTAAGGT 1629
QY 1141 gtgctgcaagtgactcatctcggaagccttggaacactgtggaagcctagcttgagagt 1200
DB 1630 GTGGCTGCAAGTGACTTCATTCGAGGCTTGACACACTGTGGAACCTAGCTTGAGAGGT 1689
QY 1201 gtgttctcccaagttgatttccgaactgacagatggttcagagtggaatgagatgctc 1260
DB 1690 GTGTTCTCCCAAGGTTGATTTTCGAATCTGCAAGATGTGTCAGCGTGAATGAGATGCC 1749
QY 1261 ctgtcgaacactgggtccctctgtaaatatcacagctcaaaaggttcaagccaagtataa 1320
DB 1750 CTGCTGACAACTGGGCTCTCTTAATATATACAGCTCAAAAGTTTCAAGCCAAAGTATAA 1809
QY 1321 ttcttccaaagctcatctccagaggtacacagcagagcgaagactcagcagtttatgag 1380
DB 1810 TTCTTTTCACAAATTCATTCAGAGGTACACAGCAGAGAAAGACTTCAGAGTTATGAGC 1869
QY 1381 tctcaatgagcgaagaggtgaccaaagggaatggtactctcagaagaatggtttccatt 1440
DB 1870 TCTCATGAGCAGGAGGAGGAGCAAGGAGGAATGTTACTTCAGAGAAATGAGTTCCATT 1929
QY 1441 tgggaatcacatccacttataagcagcgtgctccggtlaagccgtggtggtatctgtgga 1500
DB 1930 TCGGACATTCATTCACCTTATACAGCTGCTCCGGTACACCTGTGGGTATCTGTGGAA 1989
|||||

Oy 1501 gccaccaggctgttatgaagacccctcgagcagtgtaacacacggctcttcgga 1560
 |||||||
 Db 1990 gccaccaggctgttatgaagacccctcgagcagtgtaacacacggctcttcgga 2049
 |||||||
 Oy 1561 ctcccatcgccaagaagcctctctgagacaggaactcttgcaagtgtaaaaaacac 1620
 |||||||
 Db 2050 ctttccatcgccaagaagcctctctgagacaggaactcttgcaagtgtaaaaaacac 2109
 |||||||
 Oy 1621 actgagcaagaatctctgaagccataacatcaatctcttgtagatgtggaatccat 1680
 |||||||
 Db 2110 actgagcaagaatctctgaagccataacatcaatctcttgtagatgtggaatccat 2169
 |||||||
 Oy 1681 ttaatacaagagatfacatccaaatcaagccctgagccaagaatctggaactcttcaa 1740
 |||||||
 Db 2170 ttaatacaagagatfacatccaaatcaagccctgagccaagaatctggaactcttcaa 2229
 |||||||
 Oy 1741 ggtaaaagcttatataactcaactcaaggaacatcccgatctacttaattgactcttgaa 1800
 |||||||
 Db 2230 ggtaaaagctttatattatcaactcaaggaacatcccgatctacttaattgactcttgaa 2289
 |||||||
 Oy 1801 cattgcccgaattgtgcaagtgctgtgactcaatcaactggaactcttgagggagact 1860
 |||||||
 Db 2290 cattgcccgaattgtgcaagtgctgtgactcaatcaactggaactcttgagggagact 2349
 |||||||
 Oy 1861 atggtctatggaagaagcctgagcaagaacaggtggaatccacatggaagaagccca 1920
 |||||||
 Db 2350 atggtctatggaagaagcctgagcaagaacaggtggaatccacatggaagaagccca 2409
 |||||||
 Oy 1921 gaaactcaatctcccaagcaggtgtatcttcttctcaacttggaagcaggaattcagg 1980
 |||||||
 Db 2410 gaaactcaatctcccaagcaggtgtatcttcttctcaacttggaagcaggaattcagg 2469
 |||||||
 Oy 1981 actctgagagtgcaactccggagattccagcaagttgataagaagaatatacatatctg 2040
 |||||||
 Db 2470 actctgagagtgcaactccggagattccagcaagttgataagaagaatatacatatctg 2529
 |||||||
 Oy 2041 gggaaaataatcaagctctgccaagcctcaagcctgcaagaataagaatgtctggtgtg 2100
 |||||||
 Db 2530 gggaaaataatcaagctctgccaagcctcaagcctgcaagaataagaatgtctggtgtg 2589
 |||||||
 Oy 2101 gctggaagcctcagtttggtctctcagcagcctgtgaagaactattctctctcagtgtaa 2160
 |||||||
 Db 2590 gctggaagcctcagtttggtctctcagcagcctgtgaagaactattctctctcagtgtaa 2649
 |||||||
 Oy 2161 gccagctccctccacatagaagatgagaggaacatcaatctgttaacaaacttgaaacc 2220
 |||||||
 Db 2650 gccagctccctccacatagaagatgagaggaacatcaatctgttaacaaacttgaaacc 2709
 |||||||
 Oy 2221 ttgaattatcagctacagaatcaacagcgtgcccgggtgtctcagcagcagcttggt 2280
 |||||||
 Db 2710 ttgaattatcagctacagaatcaacagcgtgcccgggtgtctcagcagcagcttggt 2769
 |||||||
 Oy 2281 aacttgaagaacctacaaaagctcaaaatgaatacaataagaatgaagaagaatgct 2340
 |||||||
 Db 2770 aacttgaagaacctacaaaagctcaaaatgaatacaataagaatgaagaagaatgct 2829
 |||||||
 Oy 2341 ataacaactagctggaagcctgaaaaaactggaagaatgtgttattcttcaattgacac 2400
 |||||||
 Db 2830 ataacaactagctggaagcctgaaaaaactggaagaatgtgttattcttcaattgacac 2889
 |||||||
 Oy 2401 ttgtctgacattggaaggaaggaatgataatcaatgacagctctctgtaagtgaaacctgt 2460
 |||||||
 Db 2890 ttgtctgacattggaaggaaggaatgataatgacagctctctgtaagtgaaacctgt 2949
 |||||||
 Oy 2461 gaccttgaagaatcaatagctgctcctgctgtgtctgtaacatgcaagtgcaagtgaaatcc 2520
 |||||||
 Db 2950 gaccttgaagaatcaatagctgctcctgctgtgtctgtaacatgcaagtgcaagtgaaatcc 3009
 |||||||
 Oy 2521 gctcagaactctcacaattgtgtcacaactgagcattcttgattatcagaataatcactg 2580
 |||||||
 Db 3010 gctcagaactctcacaattgtgtcacaactgagcattcttgattatcagaataatcactg 3069
 |||||||

Oy 2581 gaaaaagatggaagaatggaagcctctcatgactgacatgcaagaatgaagcctgtaagaag 2640
 |||||||
 Db 3070 gaaaaagatggaagaatggaagcctctctcatgactgacatgcaagaatgaagcctgtaagaag 3129
 |||||||
 Oy 2641 ctccagcagatgagctgctccctgagcctgtgacgtgcaagcagcctgagcagcctgtg 2700
 |||||||
 Db 3130 ctccagcagatgagctgctccctgagcctgtgacgtgcaagcagcctgagcagcctgtg 3189
 |||||||
 Oy 2701 aaacattggaaggaagtcctcccaactcgtcaagccttggttgtaaaaaacttgagatcaca 2760
 |||||||
 Db 3190 aaacattggaaggaagtcctcccaactcgtcaagccttggttgtaaaaaacttgagatcaca 3249
 |||||||
 Oy 2761 gatcacagattagaatttgatggtcatttttgtaagaagaaccttgaaacttcag 2820
 |||||||
 Db 3250 gatcacagattagaatttgatggtcatttttgtaagaagaaccttgaaacttcag 3309
 |||||||
 Oy 2821 cagttgaaattgagcggaaatcgtgtgagcagtgatggaatgagctgtccctcaatggtgta 2880
 |||||||
 Db 3310 cagttgaaattgagcggaaatcgtgtgagcagtgatggaatgagctgtccctcaatggtgta 3369
 |||||||
 Oy 2881 ttgagaactcctaagcaatgagtggttttgacttagtaactaaagaatctcactgact 2940
 |||||||
 Db 3370 ttgagaactcctaagcaatgagtggttttgacttagtaactaaagaatctcactgact 3429
 |||||||
 Oy 2941 ccagcattagtcagaagaacttagccaagtgatattatccaaagttaactttctgcaagaact 3000
 |||||||
 Db 3430 ccagcattagtcagaagaacttagccaagtgatattatccaaagttaactttctgcaagaact 3489
 |||||||
 Oy 3001 aggcctgtgtggtggaatttgatgatatgatcctcagtgtaattatcagaagtgctttaa 3060
 |||||||
 Db 3490 aggcctgtgtggtggaatttgatgatatgatcctcagtgtaattatcagaagtgctttaa 3549
 |||||||
 Oy 3061 ctagtactgct 3072
 |||||||
 Db 3550 ctagtactgct 3561
 |||||||

RESULT 5
 AC010968
 LOCUS AC010968 160583 bp DNA HMG 18-AUG-2000
 DEFINITION Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE,
 11 unordered pieces.
 AC010968
 AC010968.5 GI:9845170
 VERSION
 HMG: HMGCS_PHASE1: HMGCS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 160583)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT
 On Aug 18, 2000 this sequence version replaced gi:8439959.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 ----- Project Information -----
 Center project name: H.NH0093002
 ----- Summary Statistics -----
 Sequencing vector: M13; 598
 Sequencing vector: plasmid; 418
 Chemistry: Dye-terminator Big Dye; 53% of reads
 Chemistry: Dye-terminator Big Dye; 53% of reads
 Assembly program: Phrap; version 0.990119

Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 159583; sum-of-contigs
Quality coverage: 6.64 in Q20 bases; agarose-fp
Quality coverage: 6.38 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1080: contig of 1080 bp in length
* 1081 1180: gap of unknown length
* 1181 2476: contig of 1296 bp in length
* 2477 2576: gap of unknown length
* 2577 5035: contig of 2459 bp in length
* 5036 5135: gap of unknown length
* 5136 9606: contig of 4471 bp in length
* 9607 17078: gap of unknown length
* 17079 17178: contig of 7372 bp in length
* 17179 27158: contig of 9980 bp in length
* 27159 27259: gap of unknown length
* 27259 45137: contig of 17879 bp in length
* 45138 45237: gap of unknown length
* 45238 65522: contig of 20285 bp in length
* 65523 65622: gap of unknown length
* 65623 91498: contig of 25876 bp in length
* 91499 91598: gap of unknown length
* 91599 116835: contig of 25237 bp in length
* 116836 160583: gap of unknown length
* 160583 160583: contig of 43648 bp in length.

location/Qualifiers
1. 160583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-9302"
1. 1080
/note="assembly_name:Contig9"
1181. 2476
/note="assembly_name:Contig12"
2577. 5035
/note="assembly_name:Contig13"
5136. 9606
/note="assembly_name:Contig14"
9707. 17078
/note="assembly_name:Contig15"
17179. 27158
/note="assembly_name:Contig16"
27259. 45137
/note="assembly_name:Contig17"
45238. 65522
/note="assembly_name:Contig18
clone_end:T7
vector_side:right"
65623. 91498
/note="assembly_name:Contig19"
91599. 116835
/note="assembly_name:Contig20
clone_end:SP6
vector_side:right"
116936. 160583
/note="assembly_name:Contig21"

BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
ORIGIN

Query Match 63.3%; Score 1946; DB 2; Length 160583;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1996; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 263 gtcttttcacagactcgaaggaaggaacttgacggtttgctcagatlttaagact 332
|||||
Db 49188 GTCCTTTTCATCAACATTCAGAAAGAGACATGACGATTTGGCTCAGATTTAAAGACT 49247

Qy 323 tglacatacccccctcttcgaactttatcccttggtgaagatgaacttattc 382
|||||
Db 49248 TGTACCATACCCCATCTTTCTGACTTTATCCCTTGGTGAAGATATGACATTTT 49307

Qy 383 ttaacttgaaaagaccttcaacagaacctgcttgyaggaaggaacacacatcacc 442
|||||
Db 49308 TTAACCTGAAAAGACCTTCAACAGAACCTGTCCGTGTGAGGAAGAGACCAACATCACC 49367

Qy 443 gcgtgagagagccttgacccctgaatgagctctctcgaagcttctcagagccctgcatactg 502
|||||
Db 49368 GCGTGAGAGACATGACCTGAATGCTCCTGACGCTCTTCAAGCCCTGCATCATTTG 49427

Qy 503 aagtggaatctggcaaggaagtcacactctgctgagcgcatgtccttgggct 562
|||||
Db 49428 AAGGGAATCTGGCAAGGCAAGTCACCTGCTGACGCGCATTTGCCATGCTCTGGGGCT 49487

Qy 563 ccggaagatgcaaggtctgacccaagtccaattgctctctcctccgtctcaagag 622
|||||
Db 49488 CCGGAAGATGCAAGGCTCTGACCAAGTTCAAAATTCGCTCTTCTCCGCTCAGCGAGG 49547

Qy 623 cccgaggtgacatttgaaacccctgtgatcaatcccggtatataccgtggaacatca 682
|||||
Db 49548 CCCGAGGTGACATTTTGAACCCCTGTGATCAACTCTGGAATATCGTGGCAATCA 49607

Qy 683 ggaagcagacatltcatgagcactgctgtaagctgagcgaaggtctcttcctctg 742
|||||
Db 49608 GGAAGCAGACATTCATGCGCATGCTGTGAAGCTGCGGAGAGGGTCTTTCTCTTTG 49667

Qy 743 atggtctcaatgaatltcaagccccaagactgcccagaatcogaagccctgataagaaa 802
|||||
Db 49668 ATGGCTCAATGAAATTCAGCCCGAGAACTCCAGAAATCGAAGCCCTGATTAAGGAAA 49727

Qy 803 accacgcttcaagaatactgtaactgtaacactacactacactgagctgaagcacatc 862
|||||
Db 49728 ACCACGCTTCAAGAAATGATGCTATGCTCACCTTACCATGATGATGCTGACCATAC 49787

Qy 863 ggcagtttgctgcccctgactgctgtaggttgaggatgatatgacaagaagagcccaagctc 922
|||||
Db 49788 GGCAGTTTGGTGCCCTGACTGCTGAGTGGGGGATATGACAGAAAGAGCCCAAGCTC 49847

Qy 923 tcatcgagaagatgctgatacgaagagctgctgaaggtgtgtgtcctaattcagaat 962
|||||
Db 49848 TCATCCGAGAAAGTGTGATGATGAGAGCTTCTGAAGGCTTGTGCTTCAAAATTCAGAAAT 49907

Qy 983 ccaagtgcttgaaggaatctcatgaagaccctcctctgtgtgtaactactgtgcaatc 1042
|||||
Db 49908 CCAAGTGTGAGGAATCTATGATGAAGACCCCTCTCTTGTGTGATCATCTTGCAATTC 49967

Qy 1043 agatgggtgaagatgagttcactctcacacacaaacaaagctgttccalaacttcatg 1102
|||||
Db 49968 AGATGGGTGAAGAGTGTGACTCTCACACACAAACAAAGCTGTTCATACCTTATG 50027

Qy 1103 atctgttgatacagaataaacaacacacataaaggtgtgtgtctgcaagtgacttcaatc 1162
|||||
Db 50028 ATCTGTGATTAACAAAAAACAACAAATTAAGGTGTGCTGCAAGTGACATTCATTC 50087

Qy 1163 ggaagcctggaacactgtggaagacctgaactctgagaggtgtgtctcccaaaattgact 1222
|||||
Db 50088 GGAAGCTGGAACACTGTGGAGACTTACTGTGAAGGTGTGTCTCCCAAGTTGATTT 50147

Qy 1223 tcaagctcagagatgtgtcagcgtgaaatgaagatgtctcgtctgacaactggtgtcctc 1282
|||||
Db 50148 TCGAACTGCAAGATGTGTCCAGGCTGAATGAGATGTGCTGACAACTGGGCTCCTCT 50207

Qy 1283 gtaaatatagctcaaaagttcaagccaaagtataaatctcttcaagtaattccagg 1342

Db	50208	GTAAATATATACAGCTCAAAAGTTCAAGCCAAATATATAATTTCTTCCAAAGTATATTCGAG	50267
OY	1343	agtacacagcagagccagaaagactcagcaglttatlgacgtlcatlgagccagaagatgta	1402
Db	50268	AGTACACAGCAGAGAGCAAGAGACTCACAGTTTATTTAGACGCTCATGAGCCAGAGAGGTGA	50327
OY	1403	ccaaggggaatggttactctgcagaaaatggttccattttcgaaacttcatccacttata	1462
Db	50328	CCAAAGGGGAATGTTACTGTTCGAAAAATGTTTCCATTTCGGACATTTACATCCACTTATA	50387
OY	1463	gcagcctgcctccggtacacctgltgggtcatcctgttgaaagccacagagcgtgttatgaaagc	1522
Db	50388	GCAGCCTGCTCCGCGTACACCTGTGGGTCACTGTGGAAAGCCACAGGGGCTGTATGAAGC	50447
OY	1523	acctgcagcagtgatatacaacagcgtcgtccttcggacttccatcgccaagagcctc	1582
Db	50448	ACCTGCAGCAGACTGTATATACACCGCTGCTGCTTCGGATTTCATTCGCAAGAGGCGTCC	50507
OY	1583	tcctggagacagaatcctttgcgaagtgtaaaaaacacacctgagcaagaatctcgaag	1642
Db	50508	TCTGGAGACAGAGAACTTTTGGCAAAGTGTGMAAAACACACTGACGACAGAAATTTCTGAAG	50567
OY	1643	ccataaacatcaatctccttctgtagagtgtygcatccattatataaagagtatacca	1702
Db	50568	CCATTAACATCAATTCCTTTGTAGAGTGTGGCATCCATTATATCAAGAGATCATCA	50627
OY	1703	aatcagccctcgaagcaagaattgtagcttcttccaagtgtaaaagcttatataccaact	1762
Db	50628	AATCAGCCCTGAGCCCAAGAAATTTGAAGCTTTCTTCAAGTAAAGCTTATATATCAAT	50687
OY	1763	caggagaaacatcccgcatctactatctgaactctcttgaaacattgcgcaattgtcgaatg	1822
Db	50688	CAGGAAACATCCCGCATTTACTTACTTATTTGAAGCTTTGAACATTTGGCCAAATGTGCAAGTG	50747
OY	1823	ctctggaacttcaataacttgacacttltatggggagcatalggtctcatgvggaaaagctg	1882
Db	50748	CTCTGGACTTCAATTAATACTGGACTTTTATGTGGGAGCTATGTGCTTCATGGGAAAAAGCTG	50807
OY	1883	cagaagacacagtgtaatccacatggaagagagcccaagaacccacatctccagcaggg	1942
Db	50808	CAGAGACACAGAGTGTGATCCACATGGAAGAGGCCCAAGAAACCTACATTTCCACAGAGG	50867
OY	1943	ctgtatcttgtlcttccaactggaagcaggaattcaagagctctggaagtgacactccggg	2002
Db	50868	CTGTATCTTTGTTCTTCAACTGTGAAGCAGGAATTCAGAGCTCTGAGAGTCACACTCCGGG	50927
OY	2003	atttcaagaagtltgaatagaagaatatacacatctctggggaaaatatltcagctgtcca	2062
Db	50928	ATTTTACGAAGTTGAATTAAGCAAGATATTCAGATATCTGTGGGAAAAATTTTACCTCTGCCA	50987
OY	2063	caagcctcaagcgtgcaaaataaagagatgctgcgtgtgtgcctbgaagcctcaattgtccc	2122
Db	50988	CAAGCCTCAGCGTGCATAAATAAAGAGATGTGCTGGTGTGGCTGCAAGCGCTCAATTTGGTCC	51047
OY	2123	tcagaacctgttaagaacatttatctctccaatggtyggaagccagttccctcacatagaag	2182
Db	51048	TCAGCACTGTGAAGAACTTTATTTCTTCATGTGTGAAGCCAGTCCCTCCACCAATAGAAG	51107
OY	2183	atgagaggcacatcacatctgttaacaaacccbtaaaaccttgatatacagactcacaga	2242
Db	51108	ATGAGAGGCACATCACTCTGTAAACAAACCTTGAATTAATTCATGACCTACAGTA	51167
OY	2243	atcaagagctgcgggt 2255	
Db	51168	ATCAACGGCTGCGCGGT 51184	

RESULT	6		
CNS0IDS3/c			
LOCUS	CNS0IDS3	138909 bp	DNA
DEFINITION	BAC sequence from the SPG4 candidate region at 2p21-p22 BAC 164M9 of C11B_978-SKB library from chromosome 2 of Homo sapiens (Human),		
		PRI	02-MAR-2000

Accession Version Keywords Source Organism	Reference Authors	Title
complete sequence. AL121653 AL121653.2 GI:7159616 HTG: HTGS_DRAFT; SPG4 genomic DNA interval. human. Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 138909)	Hazan, J., Fomknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Winkler, P., Brothier, P., Catolico, L., Barbe, Y., Burgunder, J.M., Prud'Homme, J.F., Brice, A., Fontaine, B., Hellig, R. and Weissenbach, J.	Spastin, a novel AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia Nat. Genet. (1999) in press 2 (bases 1 to 138909) Genoscope.
Direct Submission Submitted (29-FEB-2000) to the EMBL/Genbank/DBJ databases On Mar 6, 2000 this sequence version replaced gi:6002386. Location/Qualifiers 1..138909		

BASE COUNT	39243 a	28424 c	29121 g	42121 t
ORIGIN				

Query Match	60.0%;	Score 1844;	DB 9;	Length 138909;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1994; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

QY 263 gtcttttcacagacatcagaagagacttgcagcatctgtgccagatttaaagact 322

Db 75309 GGTCTTTTCATCAGACATCAGAGAGAGACTTGGACGATTTGGCTCAGGATTTAAAGGACT 75250

Db 75249 TGAACCAATACCCCACTTTCTTGACTTTATCCCTTGGTGAAGATATTTGACATTATTT 75190

363 ttacttgaaaagacaccttcacagaaacctgtctctgtggaagaaaccaacacatacc 442

Db 75189 TTAACTGTGAAGAAGCACCCTTCACAGAACCTGTCTGTGGAGGAAGSACCAACACCATCACC 75130

Db 75129 GCCTGGAGAGCGTGAACCCGAAATGSSCTTCCTGAGGGCTTCAGAGCCCTGCATCATTTG 75070

503 aagggaatctgcaagcaagcactctgctgcagcgcatgcatgctctgggt 562

Db 75069 AAGGGAACTCTGGCAAGGCAAGTCCACTCTGCTGCAGGGAATTGCCCATTGCTCTGGGGCT 75010

Df 75009 CCGGAAAGGCACGGCTTCGACCAGTTCAAAATTCGTCTTTCTTCCTCCGCTCAGCAGGG 74950
05 ccggaaadlvgcaagggccctgaccaglttcaadalcttcgtcctctctccgtccgcacagg 622

623 ccacgggtgacttttgaacctctgtatcaactcctgatatactctgcacatca 682

Db 7494.9 CCCAGGGTGGACITTTTGAACCCCTGTGATCAACTCCTGGATATATACCTGGCAATCA 74890

DB 74880 GGAAGCAGACATTCATTGGCCATGCTGTCAAGCGGCCGCGAAGAAGTTCCTTTCTCTTTC 74890
 |||||
 YY -bb3 ggaagcagaacatcatggtccatgcctgcctcaagcgcgcgagaaagggttcttttcctcttg /42

743 atggtacaatgaattcaagcccagaactgcccagaatcgaagccctgataaagaaa 802

Db 74829 ATGCGCTACAAATGAATTCAAGCCCCGAGAACTGCGCCAGAATACTGAAGCCCCGATATAAGGAAA 74770

QY 803 accacgcctcaagaacatggtcatcgtccacactacgaagtgctcgagacatac 862
 Db 74769 ACCACGCGCTTCAAGAAATGGTATCGTCACTACCTAGTGAAGGCGCATAC 74710
 QY 863 ggcgaatttgccctgacgcctgagtgaggatgatgaagaagaagcccaagctc 922
 Db 74709 GGCAGATTGGTCCCTGACGTGAGGTGGGATATGACAGAAACAGCGCCCAAGCTC 74650
 QY 923 tcatccgaagaagtcgtatcaaaagagctgtcgaagctgtgtgtccaaatccaagaat 982
 Db 74649 TCATCCGAGAAAGTGCATGATCAAGAGAGCTTGCTGAAGCTTGTTGCTCAATTCAGAAAT 74590
 QY 983 ccaagtgcttgaggaatctcatgaagaaccctcctctgtgtgtcatcacttgtaactc 1042
 Db 74589 CCAAGTGCTTGAGGAATCTCATGAAGACCCCTCTCTTGCTCATCACTTGTCATTC 74530
 QY 1043 agatgggtgaagaatgagttcactctcacacaaagaagctgtgtcatcactatg 1102
 Db 74529 AGATGGGTGAAGTGAATTCACCTCTCACACACAAACAGCGTGTCCATTCATATG 74470
 QY 1103 atctgtgtatagaagaacaaacacacataaagggtgtgtgtcgaagtgactcatc 1162
 Db 74469 ATCTGTTGATACAGAAACAAACACATAAAGGTGTGTGCTGCAAGTCACTTCAATTC 74410
 QY 1163 ggaagcctggaaccactgtggaagacctcctggaagggtgtgtgtcctccacaagttgatt 1222
 Db 74409 GGAGCCCTGACACCTCTGAGACCTAGCTCTGAGGGTGTGTCTCCACAAAGTTGATT 74350
 QY 1223 tcgaactgcaggaatgtgtccagcgttgatagaagtgtcctgctgacacactgggctcctc 1282
 Db 74349 TCGAACTGCGAGGATGTCTCCAGCTGATAGAGATGTCTGCTGACAACTGGGCTCTCT 74290
 QY 1283 gtaataataagctcaaaagttcaagccaagtaataatcttcacaagtcattccagg 1342
 Db 74289 GTAATATACAGCTCAAGAGTTCAGCCAAAGTAAATTTCTTACAAAGTCATTCACAG 74230
 QY 1343 agtacacagcagaagaagactcagcagttatctgacgtcctcatgagccagaaggatga 1402
 Db 74229 AGTACACAGCAGGAGCAAGAGCTCAGAGTTATGACGTCTCATGAGCCAGAGAGGTGA 74170
 QY 1403 ccaaggggaatgttactctcagaagaatgttccatttggaatactacacactata 1462
 Db 74169 CCAAGGGGAATGTACTTCCAGAAATGTTCATTTCCAGCAATTCACCTTATA 74110
 QY 1463 gcaagcctgcctgatacactgtggtgtcatcgttggaagccacagggcgtgtatgaagc 1522
 Db 74109 GCAGCGTGCCTCCGGTACACTGTGGGTCACTGTGGAAGCACAGGCGTGTATGAAGC 74050
 QY 1523 acctgcagcagtgatcatcaacacggctgcctctcgaacttccatctgcagaaggctc 1582
 Db 74049 ACCTGCGCAGCAGTGTATCAACACGGCGCTCTCGGACTTTCATCGCCAAAGAGCGCTC 73990
 QY 1583 tctggagacaggaatctctcagaagtgtaaaacacactgagcaagaattctgaag 1642
 Db 73989 TCTGGAGACAGGAATCTTTCAAAAGTGTGAAGAACACACTGACCAAGAAATCTGAAG 73930
 QY 1643 ccaataacacaaatccctgttagagtggtgacatccatatacaagaagatacaca 1702
 Db 73929 CCATTAACATCAATCTCTTGTAGAGTGTGCATCCATTTATATCAAGAGATACATCCA 73870
 QY 1703 aatcagcctgagccaagaatttgaagcttcttccaaggtaaaagcttatatacaact 1762
 Db 73869 AATCAGCGCCGAGCCAAAGAAATTTGAAGCTTCTTCAAGSTAAAGCTTATATCAACT 73810
 QY 1763 caagggaacatccccgatactattgacttcttgaaactttgccaattgtgcaagtg 1822
 Db 73809 CAGGGAAACATCCCGATTACTTATTGACTTCTTGAACAATTTGCCCAATTTGCAAGTG 73750
 QY 1823 ccttggaactcatataactggaacttatgtggaagctatgagcttcaaggaaagctg 1882
 Db 73749 CCTTGGACTTCATTAACGTGACTTTTATGGGGAGCTATGGCTTCATGGGAAAGAGCTG 73690
 QY 1883 cagaagaacacagtggaattccacatggaagaagccccaagaactacatccagcaggg 1942

Db 73689 CAGAAAGACACAGGTGGAATCCACATGGAAGAGGCCCCAGAAACCTACATTCACAGCAGG 73630
 QY 1943 ctgtatcttgtctctcaacttggaagaaggaattcagaactctggaagtcacactcggg 2002
 Db 73629 CTGTATCTTTGTTCTTCAACTGGAAGCAGAAATTCAGAGCTCTGAGAGTCAACTCCGG 73570
 QY 2003 attcagaagcttgaataagaagataatcatactcgggggaaataatcagctctgcca 2062
 Db 73669 ATTTACCAAGTGTGAATAGCAAGATATCAAGATATCTGGGAAATATTCAGCTCTGCCA 73510
 QY 2063 caagcctcagcctgcaaaataagaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2122
 Db 73509 CAAGCCCTCAGCGCTGCAATGAAGAGATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 73450
 QY 2123 tcgaagccttgaagaattatctctcactggttggaagccagctccctcaccataagaag 2182
 Db 73449 TCAGCACCCTGAAGACATTTATTTCTTCATGTGTGGAAGCCAGTCCCTCATAGAG 73390
 QY 2183 atgagagcacatcacatctcgtlaacaacctgaaaccttgaglatcatgacctacaga 2242
 Db 73389 ATGAGAGGCACATCATCATCTGAACAACCTGAAGAACTTGAGATATCATGACCTACAGA 73330
 QY 2243 atcaagcgtcgcgggt 2259
 Db 73329 ATCAACGCGTGCAGG 73313
 RESULT 7
 IR2005417
 LOCUS IR2005417 1355 bp mRNA PRI 16-JUL-2000
 DEFINITION Homo sapiens mRNA full length insert cDNA clone EUOIMAGE 2005417.
 ACCESSION AL389934
 VERSION AL389934.1 GI:9367839
 KEYWORDS FLI.CDNA.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1355)
 AUTHORS Aufray C., Ansojorge, W., Ballabio, A., Estvill, X., Gibson, K.,
 Leirach, H., Ponsla, A., and Lundberg, J.
 TITLE The European IMAGE consortium for integrated Molecular analysis of
 human gene transcripts
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1355)
 TITLE Direct Submission
 JOURNAL Submitted (15-JUL-2000) Dept. Genetica Molecular, Institut de
 Recerca Oncologica (IRO), Hospital Duran i Reynals, Av. Gran Via
 s/n Km 2,7 U/Hospital de Llobregat, 08907 Barcelona, Catalunya,
 SPAIN. Tel: ++34-93-260-7775 Fax: ++34-93-260-7776 WWW site:
 http://www.iro.es e-mail enquiries: lsunoy@iro.es
 COMMENT EURO-IMAGE Consortium Contact: Aufray C
 CNRS UPR 420 - Genetique Molculaire et Biologie du Developpement
 IFR 1221 - Rue Guy Moquet 19, Batiment G - BP 8
 94801 Villeneuve Cedex, FRANCE
 Tel: ++33-1-49 58 34 98
 Fax: ++33-1-49 58 35 09
 e-mail: aufray@infobiogen.fr
 This clone is available royalty-free through IMAGE Consortium
 distributors.
 IMPORTANT: This sequence represents the full insert of this IMAGE
 cDNA clone. No attempt has been made to verify whether this
 corresponds to the full length of the original mRNA from which it
 was derived.
 FEATURES
 Location/Qualifiers
 1..1355
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="IMAGE cDNA clone 2005417"
 /clone_1ib="NCI_CGAP_Pan1"

Query Match	Best local similarity	Score	DB	Length
2258	97	100.0%	9	1395
539	97	100.0%	9	1395
2318	97	100.0%	9	1395
599	97	100.0%	9	1395
2378	97	100.0%	9	1395
659	97	100.0%	9	1395
2438	97	100.0%	9	1395
719	97	100.0%	9	1395
2458	97	100.0%	9	1395
779	97	100.0%	9	1395
2558	97	100.0%	9	1395
839	97	100.0%	9	1395
2618	97	100.0%	9	1395
899	97	100.0%	9	1395
2678	97	100.0%	9	1395
959	97	100.0%	9	1395
2738	97	100.0%	9	1395
1019	97	100.0%	9	1395
2798	97	100.0%	9	1395
1079	97	100.0%	9	1395
2858	97	100.0%	9	1395
1139	97	100.0%	9	1395
2918	97	100.0%	9	1395
1199	97	100.0%	9	1395
2978	97	100.0%	9	1395
1259	97	100.0%	9	1395
3038	97	100.0%	9	1395

DB	1319	GTGTTATTACAGCTGCTTTTAACAGTAAGTACGT	1353
RESULT	9		
LOCUS	AY027789	768 bp	mRNA
DEFINITION	Homo sapiens CLANC (CLAN1) mRNA, complete cds.		PRI
ACCESSION	AY027789		
VERSION	AY027789.1	GI:14324116	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 768)		
TITLE	Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.		
JOURNAL	Genomics. 75 (1-3), 77-83 (2001)		
MEDLINE	21365712		
PUBMED	11472070		
REFERENCE	2 (bases 1 to 768)		
AUTHORS	Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-FEB-2001) Program on Apoptosis and Cell Death Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES	Location/Qualifiers		
source	1..768		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="2"		
	/map="2p22-p21"		
	/tissue-type="Lung"		
gene	1..768		
	/gene="CLAN1"		
CDS	277..747		
	/gene="CLAN1"		
	/codon_start=1		
	/product="CLANC"		
	/protein_id="AAK14778.1"		
	/db_xref="GI:14324117"		
	/translation="MNFIDKNSRALIORMGNTVTKLTDDLFWNNVNNREVNITICE KVEDARGIITHMLKKSGESCNEFLKSEMNYPDLONGOSTLPHQTSBGDDLLA QDLDLHYHYSFLNFFPLGEDIDILFPLKSTFTEPYLMRKQHHHREQLTVL"		
BASE COUNT	218 a 157 c 180 g 213 t		
ORIGIN			
Query Match	15.0%: Score 461; DB 9; Length 768;		
Best Local Similarity	100.0%: Pred. NO. 1.4e-236; Mismatches 0; Gaps 0;		
Matches 461:	Conservative 0; Mismatches 0; Indels 0;		
QY	1 atgaattcacaataagacaatagccgagcccttattcaaaagtggagatgactgttata 60		
DB	277 ATGATTTTCATTAAGAGACAAATAGCCGAGCCCTTATTCAAGAAAGGGAATGACTGTTATA 336		
QY	61 aagcaatcacagatgacctattgtatggaatggttcgtaatcgcgaaagaataacatc 120		
DB	337 AAGCAAAATCAAGAGACCTATTGTATGGAATGTTGGAATCCGGAAGAATAACATC 396		
QY	121 atttcctcggaagaagtggacaagaatcctctagaagagatcatcacaatgatttgaa 180		
DB	397 ATTTCTCTCGAAGAGAGTGGACGAGATCTCTCTAGAGGATCATTCACATGATTTTGA 456		
QY	181 aagagttcaagagctctgtaacctcttcttaaatcccttaagagagtggaaactatccta 240		
DB	457 AAGGTCAGAGTCTCTGTAACCTCTTTCTTAATCCCTTAAGAGATGGAAGTATCTCTTA 516		
QY	241 tttagagacttgaataggacaaagtcttttttcaatagaacaatcaagaagagacttgaagat 300		
DB	517 TTTCAGGACTTGGAATGAGCAAAAGTCTTTTTCATCATGACATCAGAAAGGAGACTTGGAGCAT 576		

Qy 301 ttggctcagagattaaagagactgtacacatcccatcttctgaacttataccctt 360
|||||
Db 577 TTGGCTCAGAGATTAAAGACTGTACACATCCCATCTTTCTGAAGTTATCCCTT 636
|||||
Qy 361 ggtgaagatattgacatttacttgaactgaagaacaccttcagaaactgtctgtg 420
|||||
Db 637 GGTGAAGATTATGACATATTTTAACTTGAAGAAGACACCTTCACAGAACCTGCTGTGG 696
|||||
Qy 421 aggaagac 461
|||||
Db 697 AGGAGAGAC 737
|||||
RESULT 10
CNS01D58/c CNS01D58 162692 bp DNA PRI 02-MAR-2000
LOCUS BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 563N04
DEFINITION of RPEC-11 library from chromosome 2 of Homo sapiens (Human),
complete sequence.
ACCESSION AL121658.2 GI:7159619
VERSION AL121658
KEYWORDS HTG; HTGS_DRAFT; SPG4 genomic DNA interval.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 162692)
AUTHORS Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
Attiguenave,F., Davoine,C.S., Crnaud,C., Durr,A., Wincker,P.,
Brotier,P., Cattoilco,L., Barbe,V., Burgunder,J.M.,
Prud'Homme,J.F., Brice,A., Fontaine,B., Hellig,R. and
Weissenbach,J.
TITLE Spastin, a novel AAA protein, is altered in the most frequent form
of autosomal dominant spastic paraplegia
JOURNAL Nat. Genet. (1999) In press
REFERENCE 2 (bases 1 to 162692)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (29-FEB-2000) to the EMBL/Genbank/DBJ databases
COMMENT On Mar 6, 2000 this sequence version replaced gi:6002391.
FEATURES
source location/Qualifiers
1..162692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="RPEC-11"
/clone="563N04"
BASE COUNT 44981 a 32017 c 32887 g 52792 t 15 others
ORIGIN
Query Match 9.5%; Score 292; DB 9; Length 162692;
Best Local Similarity 100.0%; Pired. No. 1.6e-145;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2781 aggtgacatttttgaagaacccctgaaaaaacttcagacagtgaatttggcgga 2840
|||||
Db 152218 AGGTGACATTTTGGAAAGAACCCCTGAAAAAATTCAGAGGTGAATTTGGCGGAAA 152159
|||||
Qy 2841 tctgtgagcagtgatgagtgctgtcctcatcagtggttatttgaagaatctaaaca 2900
|||||
Db 152158 TCGGTGAGCAGTGATGATGGCTTCATGGGTATTTGGAATCTTAACAAAT 152099
|||||
Qy 2901 agtgttttctacttagtactaaagaattctactcgtatccagcatagtcagaact 2960
|||||
Db 152098 AGTGTTTTGTACTTACTTAAGAAATTTCTACCTGATCCAGCATAGTCAGAAAAT 152039
|||||
Qy 2961 tagccaagtgtatccaagttacttctcgaagaagctagcgtgtgtgtgtgcaact 3020
|||||
Db 152038 TAGCCAAGTGTATCCAGTTAACTTTCTGCAAGAAAGCTAGGCTTGGTGGCAAT 151979
|||||
Qy 3021 tgatgatgatgatcagtgcttattacagtgctttaaactagtaactgct 3072
|||||

Db 151978 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 151927
|||||
RESULT 11
AC011232/c AC011232 185281 bp DNA HTG 10-MAR-2001
LOCUS Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
7 unordered pieces.
DEFINITION AC011232.7 GI:13270720
ACCESSION AC011232
VERSION AC011232.7
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 185281)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185281)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 10, 2001 this sequence version replaced gi:9799811.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0078E13
----- Summary Statistics -----
Sequencing vector: M13; 57k
Sequencing vector: Plasmid; 40k
Chemistry: Dye-primer ET; 48% of reads
Chemistry: Dye-terminator Big Dye; 50% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182798 bases at least Q40
Consensus quality: 183538 bases at least Q30
Consensus quality: 184045 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 184681; sum-of-contigs
Quality coverage: 7.07 in Q20 bases; agarose-fp
Quality coverage: 7.33 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1985: contig of 1985 bp in length
* 1986 2085: gap of unknown length
* 2086 8183: contig of 6098 bp in length
* 8184 8284: gap of unknown length
* 8284 22741: contig of 14458 bp in length
* 22742 22841: gap of unknown length
* 22842 41446: contig of 18605 bp in length
* 41447 41547: gap of unknown length
* 41547 69182: contig of 27636 bp in length
* 69183 69282: gap of unknown length
* 69283 11231: contig of 42849 bp in length
* 11232 112231: gap of unknown length
* 112232 185281: contig of 73050 bp in length.
location/Qualifiers
1..185281
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"

```
misc_feature      /clone="RP11-78E13"
                  1..1985
misc_feature      /note="assembly_name:Contig11"
                  2086..8183
misc_feature      /note="assembly_name:Contig12"
                  8284..22741
misc_feature      /note="assembly_name:Contig13"
                  22842..41446
misc_feature      /note="assembly_name:Contig14"
                  41547..69182
misc_feature      /note="assembly_name:Contig15"
                  69283..112131
misc_feature      /note="assembly_name:Contig16"
                  11232..185281
misc_feature      /note="assembly_name:Contig17"
                  11232..185281
BASE COUNT      52719 a 36957 c 37187 g 57816 t 602 others
ORIGIN
Query Match      9.5%; Score 292; DB 2; Length 185281;
Best Local Similarity 100.0%; Pred.No. 1.6e-145;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2781 aggtgcatttttggaaagaacccctgaaacacccctcagcagttgaattggcgggaaa 2840
      |||
DB 160183 AGGTGCAATTTTGGAAAGAACCCCTCGAAAACTCCACAGTGTGAATTTGGCGGAAA 160124
      |||
QY 2841 tcgtgagacagatgagatgagcttcctcagtgatgtaattgaaatcttaagcaatt 2900
      |||
DB 160123 TCCTGTGACAGCAGATGATGATGCTTGCCTTCATGGGTGTGATTTGAGATCTTAAGCAATT 160064
      |||
QY 2901 agtgcattttgacttaagtaagaattctcactcagcagcattagtcagaacatt 2960
      |||
DB 160063 AGTGTATTTTGAATTTAGTACTAAGAAATTTTACCTGATCCAGCATTAAGTCAAGAAACT 160004
      |||
QY 2961 tagccaagtggttccaagttaacttttcgcagaagaagcctgagctgttggtggcgaatt 3020
      |||
DB 160003 TACCCAAGTGTTTCCAGATTACTTTTCGCAAGAAAGCTAGGCTTTGGTGGCAATT 159944
      |||
QY 3021 tgatgatgatcatcagtgcttattcaagtgctttaaactagtaactgct 3072
      |||
DB 159943 TGATGATGATGATCTCAGTGTATTATACAGTGCTTTTAACTGATGATCTGCT 159892
      |||

RESULT 12
LOCUS      AY027790      578 bp      mRNA      PRI      20-JUL-2001
DEFINITION Homo sapiens CLAND (CLAN1) mRNA, complete cds.
ACCESSION      AY027790
VERSION      AY027790.1 GI:14324118
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.
REFERENCE
  1 (bases 1 to 578)
AUTHORS      Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE      Clam, a novel human ced-4-like gene
JOURNAL      Genomics. 75 (1-3), 77-83 (2001).
MEDLINE      21365712
PUBMED      11472070
REFERENCE
  2 (bases 1 to 578)
AUTHORS      Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE      Direct Submission
JOURNAL      Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
  source
    1..578
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2p22-p21"
```

```
gene
  1..578
  /gene="CLAN1"
  /gene="CLAN1"
  /gene="CLAN1"
  /note="CARD protein"
  /product="CLAND"
  /product_start=1
  /protein_id="AAK14779.1"
  /db_xref="GI:14324119"
  /translation="MKFIRKDNSRALIQRMKVIVIKQIPDLEFVNVLYNREEVNIICCE
KVEQDAKRTIIMILKKSGSESCWLFKSLKEMNPLFDPLNQSLLTA"
BASE COUNT      172 a 106 c 143 g 157 t
ORIGIN
Query Match      8.8%; Score 269; DB 9; Length 578;
Best Local Similarity 100.0%; Pred.No. 4.9e-133;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaattcataagacataagcagcagccctattcaagaatgggaatgactgttata 60
      |||
DB 277 ATGAATTTCTAAAGCAGCATAGCCGAGCCCTTATTAAAGATGGGAATGACTGTATA 336
      |||
QY 61 aagcaaatcacagatgacctatttgatgaaatgctcgtatcgagaagaataacatc 120
      |||
DB 337 AAGCAATTCACAGATGACATATTGTATGATGAATGTTCTGAATCGCGAAGAAATCAATC 396
      |||
QY 121 atttgctggagaagaatggtgagcagatgctgctcagaaggagatcaatcatgatttgaaa 180
      |||
DB 397 ATTTGCTGCGAAGAGGTGAGGAGATGCTGTAGAGGATTCATTCATGATGATTTTGAAA 456
      |||
QY 181 aaggttcaagatcctcttaaccccttcttaaatcccttaaggagatgagaaactatcctta 240
      |||
DB 457 AAGGTTCAAGATCTCTTAACCTCTTTCTTAATCCCTTAAGGAGTGGAACTTCTCTTA 516
      |||
QY 241 ttccaggaactggaatggacaagaactcttt 269
      |||
DB 517 TTTCAGGACTTGAATGAGCAAAAGTCTTTT 545
      |||

RESULT 13
LOCUS      G55568/c      553 bp      DNA      STS      30-MAR-2000
DEFINITION      SHGC-100923 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION      G55568
VERSION      G55568.1 GI:6120887
KEYWORDS      STS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.
REFERENCE
  1 (bases 1 to 553)
AUTHORS      Olivier,M. and Cox,D.R.
TITLE      Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL      Unpublished (2000)
COMMENT
  Contact: Michael Olivier, David R. Cox
  Stanford Human Genome Center
  Stanford University School of Medicine
  4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
  Tel: (650) 320-5800
  Fax: (650) 320-5801
  Email: oliviereshgc.stanford.edu
  Primer A: AATAGGGGGCAAAATATAGCAAA
  Primer B: TAACACCCCTTTGTTCTTCICA
  STS size: 322
  PCR Profile:
    Initial incubation: 95 degrees C for 10 minutes
    Denaturation: 94 degrees C for 30 seconds
    Annealing: 60 degrees C for 30 seconds
    Polymerization: 72 degrees C for 23 seconds
    PCR cycles: 30
```

Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AmpliTaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
BAC ends sequenced at TIGR from the RPC11 BAC library. Designed and developed at the Stanford Human Genome Center.
FEATURES
source Location/Qualifiers
1..553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="2"
clone_id="Human"
STS
5..326
5..27
primer_bind complement(304..326)
BASE COUNT 170 a 107 c 114 g 162 t
ORIGIN

Query Match 5.5%; Score 170; DB 11; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.2e-79;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2615 tcgacagatgaacgtctctagaacagctaccgcactgagtctgcctgggctgtgacg 2674
|||||
Db 258 TCGACAGCATGACGTCTAGACACAGCTCACCGCATGATCTGCGGCGCTGTGACG 199
|||||
QY 2675 tcgacagcagctgacgcctgttgaacacattggagaggtcccaactcgtcaagc 2734
|||||
Db 198 TGCAGGCGCCTGAGCGCTGTGAAACATTTGGAGAGGTCCACACACTCGCAAGC 139
|||||

QY 2735 ttgggttgaacactgagacacacagatcacagatagatattaggt 2784
|||||
Db 138 TTGGGTGAAAACTGAGACTCACAGATACAGAGATTAGATTAGT 89
|||||

RESULT 14
AC074195/c DNA HTG 20-AUG-2000
LOCUS Homo sapiens chromosome 11 clone RP11-750A9, WORKING DRAFT
DEFINITION
SEQUENCE, 25 unordered pieces.
AC074195
AC074195 GI:9799883
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 175152)
Waterston,R.H.
The sequence of Homo sapiens clone
JOURNAL
Unpublished
2 (bases 1 to 175152)
Waterston,R.H.
Direct Submission
Submitted (16-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 13, 2000 this sequence version replaced gi:9743495.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0750A09
----- Summary Statistics -----
Sequencing vector: Plasmid; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 155181 bases at least Q40
Consensus quality: 161068 bases at least Q30
Consensus quality: 163835 bases at least Q20
Insert size: 174000; agarose-fp
Insert size: 172752; sum-of-contigs
Quality coverage: 4.16 in Q20 bases; agarose-fp
Quality coverage: 4.34 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1335: contig of 1335 bp in length
1336 1435: gap of unknown length
1436 3074: contig of 1639 bp in length
3075 3174: gap of unknown length
3175 5534: contig of 2360 bp in length
5535 5634: gap of unknown length
5635 8287: contig of 2653 bp in length
8288 8387: gap of unknown length
8388 10415: contig of 2028 bp in length
10416 10515: gap of unknown length
10516 12875: contig of 2260 bp in length
12876 12975: gap of unknown length
12976 16283: contig of 3308 bp in length
16284 16384: gap of unknown length
16385 19546: contig of 3163 bp in length
19547 19646: gap of unknown length
19647 24020: contig of 4374 bp in length
24021 24120: gap of unknown length
24121 28746: contig of 4626 bp in length
28747 28846: gap of unknown length
28847 33535: contig of 4689 bp in length
33536 33635: gap of unknown length
33636 38255: contig of 4620 bp in length
38256 38355: gap of unknown length
38356 43277: contig of 4922 bp in length
43278 43377: gap of unknown length
43378 47903: contig of 4526 bp in length
47904 48003: gap of unknown length
48004 52384: contig of 4381 bp in length
52385 52484: gap of unknown length
52485 58975: contig of 6491 bp in length
58976 59075: gap of unknown length
59076 67303: contig of 8228 bp in length
67304 67403: gap of unknown length
67404 74849: contig of 7446 bp in length
74850 74950: gap of unknown length
74951 83615: contig of 8666 bp in length
83616 83716: gap of unknown length
83717 94881: contig of 11166 bp in length
94882 94981: gap of unknown length
94982 104766: contig of 9785 bp in length
104767 104866: gap of unknown length
104867 118097: contig of 13231 bp in length
118098 118197: gap of unknown length
118198 132815: contig of 14618 bp in length
132816 132915: gap of unknown length
132916 148952: contig of 16037 bp in length
148953 149052: gap of unknown length

```
FEATURES      * 149053 175152: contig of 26100 bp in length.
source
1.175152
/organism="Homo sapiens"
/db xref="taxon:9606"
/chromosome="11"
/clone="RP11-750A9"
1.1335
misc_feature  /note="assembly_name:Contig13"
misc_feature 1436..3074
              /note="assembly_name:Contig16"
misc_feature 3175..5534
              /note="assembly_name:Contig17"
misc_feature 5635..8287
              /note="assembly_name:Contig18"
misc_feature 8388..10415
              /note="assembly_name:Contig19"
misc_feature 10516..12875
              /note="assembly_name:Contig20"
misc_feature 12976..16283
              /note="assembly_name:Contig21"
misc_feature 16384..19546
              /note="assembly_name:Contig22"
misc_feature 19647..24020
              /note="assembly_name:Contig23"
misc_feature 24121..28746
              /note="assembly_name:Contig24"
misc_feature 28847..33535
              /note="assembly_name:Contig25"
misc_feature 33636..38255
              /note="assembly_name:Contig26"
misc_feature 38356..43277
              /note="assembly_name:Contig27"
misc_feature 43378..47903
              /note="assembly_name:Contig28"
misc_feature 48004..52384
              /note="assembly_name:Contig29"
misc_feature 52485..58975
              /note="assembly_name:Contig30"
misc_feature 59076..67303
              /note="assembly_name:Contig31"
misc_feature 67404..74849
              /note="assembly_name:Contig32"
misc_feature 74950..83615
              /note="assembly_name:Contig33"
misc_feature 83716..94881
              /note="assembly_name:Contig34"
misc_feature 94982..104766
              /note="assembly_name:Contig35"
misc_feature 104867..118097
              /note="assembly_name:Contig36"
misc_feature 118198..132815
              /note="assembly_name:Contig37"
misc_feature 132916..148952
              /note="assembly_name:Contig38"
misc_feature 149053..175152
              /note="assembly_name:Contig39"
BASE COUNT 42097 a 43906 c 45616 g 41104 t 2429 others
ORIGIN
Query Match      0.7%; Score 23; DB 2; Length 175152;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1887 agacacagtggaatcacatg 1909
   |||||||
DB 80203 AGACACAGGTGGAATCCACATGG 80181

RESULT 15
AC019059/c AC019059 183556 bp DNA HTG 07-JUL-2000
```

```
DEFINITION Homo sapiens chromosome 11 clone RP11-125F14, WORKING DRAFT
SEQUENCE 31 unordered pieces.
AC019059
AC019059.4 GI:8567959
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183556)
Waterston,R.H.
REFERENCE
AUTHORS The sequence of Homo sapiens clone
TITLE Unpublished
JOURNAL 2 (bases 1 to 183556)
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 15, 2000 this sequence version replaced gi:7684541.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0125F14
----- Summary Statistics -----
Sequencing vector: M13; 70%
Sequencing vector: plasmid; 30%
Chemistry: Dye-terminator Big Dye; 30% of reads
Chemistry: Dye-terminator Big Dye; 30% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164779 bases at least Q40
Consensus quality: 170994 bases at least Q30
Consensus quality: 174798 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 180556; sum-of-ctrls
Quality coverage: 3.76 in Q20 bases; agarose-fp
Quality coverage: 3.61 in Q20 bases; sum-of-ctrls
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1181: contig of 1181 bp in length
1182 1181: gap of unknown length
1282 2783: contig of 1502 bp in length
2784 2883: gap of unknown length
2884 4618: contig of 1735 bp in length
4619 4718: gap of unknown length
4719 6318: contig of 1600 bp in length
6319 6418: gap of unknown length
6419 8952: contig of 2534 bp in length
8953 9052: gap of unknown length
9053 11436: contig of 2384 bp in length
11437 11536: gap of unknown length
11537 14108: contig of 2572 bp in length
14109 14208: gap of unknown length
14209 16184: contig of 1976 bp in length
16185 16284: gap of unknown length
16285 20536: contig of 4252 bp in length
20537 20636: gap of unknown length
20637 25264: contig of 4628 bp in length
25265 25364: gap of unknown length
25365 29797: contig of 4433 bp in length
29798 29897: gap of unknown length
29898 33626: contig of 3729 bp in length
33627 33726: gap of unknown length
```

```

* 37327 37341: contig of 3615 bp in length
* 37342 37441: gap of unknown length
* 37442 41642: contig of 4201 bp in length
* 41643 41742: gap of unknown length
* 41743 47291: contig of 5549 bp in length
* 47292 47391: gap of unknown length
* 47392 50826: contig of 3435 bp in length
* 50826 50926: gap of unknown length
* 50927 58180: contig of 7254 bp in length
* 58181 58280: gap of unknown length
* 58281 63165: contig of 4885 bp in length
* 63166 63265: gap of unknown length
* 63266 68481: contig of 5216 bp in length
* 68482 73841: contig of 5260 bp in length
* 73842 73941: gap of unknown length
* 73942 79920: contig of 5978 bp in length
* 79920 80020: gap of unknown length
* 80020 86722: contig of 6702 bp in length
* 86722 86821: gap of unknown length
* 86822 92876: contig of 6055 bp in length
* 92877 92976: gap of unknown length
* 92977 98096: contig of 5120 bp in length
* 98097 98196: gap of unknown length
* 98197 106663: contig of 8467 bp in length
* 106664 106763: gap of unknown length
* 106764 117153: contig of 10390 bp in length
* 117154 117253: gap of unknown length
* 117254 128332: contig of 11079 bp in length
* 128333 128432: gap of unknown length
* 128433 139368: contig of 10936 bp in length
* 139369 139468: gap of unknown length
* 139469 154140: contig of 14672 bp in length
* 154141 154240: gap of unknown length
* 154241 167723: contig of 13483 bp in length
* 167724 167824: gap of unknown length
* 167824 183556: contig of 15733 bp in length.
Location/Qualifiers
1. 183556

```

```

FEATURES
source
1. 183556
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-125F14"

```

```

misc_feature /note="assembly_name:Contig9"
1282. 2783
/misc_feature /note="assembly_name:Contig8"
2884. 4618
/misc_feature /note="assembly_name:Contig9"
4719. 6318
/misc_feature /note="assembly_name:Contig10"
6419. 8952
/misc_feature /note="assembly_name:Contig11"
9053. 11436
/misc_feature /note="assembly_name:Contig12"
11537. 14108
/misc_feature /note="assembly_name:Contig13"
14209. 16184
/misc_feature /note="assembly_name:Contig14"
16285. 20536
/misc_feature /note="assembly_name:Contig15"
20637. 25264
/misc_feature /note="assembly_name:Contig16"
25365. 29797
/misc_feature /note="assembly_name:Contig17"
29898. 33626
/misc_feature /note="assembly_name:Contig18"
33727. 37341
/misc_feature /note="assembly_name:Contig19"
37442. 41642
/misc_feature /note="assembly_name:Contig20"
41743. 47291
/misc_feature /note="assembly_name:Contig21"

```

```

misc_feature 47392. 50826
/note="assembly_name:Contig22"
misc_feature 50927. 58180
/note="assembly_name:Contig23"
misc_feature 58281. 63165
/note="assembly_name:Contig24"
misc_feature 63266. 68481
/note="assembly_name:Contig25"
misc_feature 68582. 73841
/note="assembly_name:Contig26"
misc_feature 73942. 79919
/note="assembly_name:Contig27"
misc_feature 80020. 86721
/note="assembly_name:Contig28"
misc_feature 86822. 92876
/note="assembly_name:Contig29"
misc_feature 92977. 98096
/note="assembly_name:Contig30"
misc_feature 98197. 106663
/note="assembly_name:Contig31"
misc_feature 106764. 117153
/note="assembly_name:Contig32"
misc_feature 117254. 128332
/note="assembly_name:Contig33
clone_end:SP6
vector_side:left"
misc_feature 128433. 139368
/note="assembly_name:Contig34
clone_end:T7
vector_side:right"
misc_feature 139469. 154140
/note="assembly_name:Contig35"
misc_feature 154241. 167723
/note="assembly_name:Contig36"
misc_feature 167824. 183556
/note="assembly_name:Contig37"
BASE COUNT 43865 a 46765 c 45344 g 44562 t 3020 others
ORIGIN

```

```

Query Match 0.7%; Score 23; DB 2; Length 183556;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1887 agaccacgtgtggaaccacatg 1909
|||||
DB 103946 AGACACAGTGTGAATCCACATGG 103924

```

```

RESULT 16
AC090582 188459 bp DNA HTG 04-JUL-2001
LOCUS Homo sapiens chromosome 11 clone RP11-125F14 map 11, WORKING DRAFT
DEFINITION AC090582
SEQUENCE 15 unordered pieces.
AC090582
VERSION AC090582.3 GI:14595831
KEYWORDS HTG; HTGS_Phrase1; HTGS_DRAFT; HTGS_Fulltop.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188459)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-125F14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188459)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Fairo,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

```

TITLE
JOURNAL

COMMENT

Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menes, L., Mihova, T., Mlangi, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Souarez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced g1:13605991.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L11783

Center clone name: 125_F_14

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182057 bases at least Q40
Consensus quality: 184539 bases at least Q30
Consensus quality: 185578 bases at least Q20
Insert size: 192000; agarose-1p
Insert size: 187059; sum-of-ctrls
Quality coverage: 8.4 in Q20 bases; sum-of-ctrls
Quality coverage: 8.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5671: contig of 5671 bp in length
* 5672 5771: gap of 100 bp
* 5772 6377: contig of 606 bp in length
* 6378 6477: gap of 100 bp
* 6478 7122: contig of 645 bp in length
* 7123 7222: gap of 100 bp
* 7223 8771: contig of 1549 bp in length
* 8772 8871: gap of 100 bp
* 8872 10666: contig of 1795 bp in length
* 10667 10766: gap of 100 bp
* 10767 13839: contig of 3073 bp in length
* 13840 13939: gap of 100 bp
* 13940 18026: contig of 4087 bp in length
* 18027 18126: gap of 100 bp
* 18127 24163: contig of 6037 bp in length
* 24164 24263: gap of 100 bp
* 24264 31828: contig of 7565 bp in length
* 31829 31928: gap of 100 bp
* 31929 69676: contig of 37748 bp in length
* 69677 69776: gap of 100 bp
* 69777 90436: contig of 20660 bp in length
* 90437 90536: gap of 100 bp
* 90537 117488: contig of 26952 bp in length
* 117489 117588: gap of 100 bp

* 117589 145954: contig of 28366 bp in length
* 145955 146054: gap of 100 bp
* 146055 174878: contig of 28824 bp in length
* 174879 174978: gap of 100 bp
* 174979 188459: contig of 13481 bp in length.

Location/Qualifiers

source

1. 188459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-125F14"
/clone_lib="RPCI-11 Human Male BAC"
1. 5671
/note="assembly_fragment
clone_end:SP6
vector_side:left"

misc_feature
5772..6377
/note="assembly_fragment"
6478..7122
/note="assembly_fragment"
7223..8771
/note="assembly_fragment"
8872..10666
/note="assembly_fragment"
10767..13839
/note="assembly_fragment"
13940..18026
/note="assembly_fragment"
18127..24163
/note="assembly_fragment"
24264..31828
/note="assembly_fragment"
31929..69676
/note="assembly_fragment"
69777..90436
/note="assembly_fragment"
90537..117488
/note="assembly_fragment"
117589..145954
/note="assembly_fragment"
146055..174878
/note="assembly_fragment"
174979..188459
/note="assembly_fragment
clone_end:T7
vector_side:right"

BASE COUNT 45628 a 47237 c 48121 g 46068 t 1405 others
ORIGIN

Query Match 0.7%; Score 23; DB 2; Length 188459;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1887 agacacagctggaatccacatg 1909
|||||
Db 85215 AGACACAGTGGATCCACATGG 85237

RESULT 17

AC090559

LOCUS

AC090559 204487 bp DNA HTG 11-JUL-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-750H9 map 11, WORKING DRAFT
SEQUENCE 16 unordered pieces.

AC090559

VERSION

AC090559.3

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AC090559 204487 bp DNA HTG 11-JUL-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-750H9 map 11, WORKING DRAFT
SEQUENCE 16 unordered pieces.
AC090559.3 GI:14670098
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 204487)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-750H9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204487)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouknight,B., Brown,A.,
Camataia,J., Campiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,W., Grand-Pierre,N.,
Hagos,B., Heath,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karates,A., LaRoque,K., Lamazares,R., Landers,T.,
Lhocck,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punukhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Strange-Rhmann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

COMMENT

Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 11, 2001 this sequence version replaced g1:14150930.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: l11075

Center clone name: 750_H_9

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 199468 bases at least Q40

Consensus quality: 201173 bases at least Q30

Consensus quality: 202030 bases at least Q20

Insert size: 200000; agarose-fp

Insert size: 202987; sum-of-ctnigs

Quality coverage: 10.6 in Q20 bases; agarose-fp

Quality coverage: 10.6 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 450: contig of 450 bp in length
* 451 550: gap of 100 bp
* 551 3157: contig of 2607 bp in length
* 3158 3257: gap of 100 bp
* 3258 4660: contig of 1403 bp in length
* 4661 4760: gap of 100 bp
* 4761 6875: contig of 2115 bp in length
* 6876 6975: gap of 100 bp
* 6976 9750: contig of 2775 bp in length
* 9751 9850: gap of 100 bp
* 9851 12616: contig of 2766 bp in length
* 12617 12716: gap of 100 bp
* 12717 21088: contig of 8372 bp in length

* 21089 21188: gap of 100 bp
* 21189 27401: contig of 6213 bp in length
* 27402 27501: gap of 100 bp
* 27502 35375: contig of 7874 bp in length
* 35376 35475: gap of 100 bp
* 35476 77415: contig of 41940 bp in length
* 77416 77515: gap of 100 bp
* 77516 91045: contig of 13530 bp in length
* 91046 91145: gap of 100 bp
* 91146 113825: contig of 22680 bp in length
* 113826 113925: gap of 100 bp
* 113926 137980: contig of 24055 bp in length
* 137981 138080: gap of 100 bp
* 138081 158403: contig of 20323 bp in length
* 158404 158503: gap of 100 bp
* 158504 194539: contig of 36036 bp in length
* 194540 194639: gap of 100 bp
* 194640 204487: contig of 9848 bp in length.
Location/Qualifiers
1. 204487

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone.lib="RP11-750H9"

/clone.lib="RP11-750H9"

1. 450

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

551. 3157

/note="assembly_fragment"

3258. 4660

/note="assembly_fragment"

4761. 6875

/note="assembly_fragment"

6976. 9750

/note="assembly_fragment"

9851. 12616

/note="assembly_fragment"

12717. 21088

/note="assembly_fragment"

21189. 27401

/note="assembly_fragment"

27502. 35375

/note="assembly_fragment"

35476. 77415

/note="assembly_fragment"

77516. 91045

/note="assembly_fragment"

91146. 113825

/note="assembly_fragment"

113926. 137980

/note="assembly_fragment"

138081. 158403

/note="assembly_fragment"

158504. 194539

/note="assembly_fragment"

194640. 204487

/note="assembly_fragment"

clone_end:T7

vector_side:right"

BASE COUNT 52691 a 48290 c 49218 g 52785 t 1503 others

ORIGIN

Query Match 0.7%; Score 23; DB 2; Length 204487;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1887 agacacagtggaatccacatg 1909

|||||

Db 72357 AGACACAGTGCATCCATG 72379

```

RESULT 18
AP002957 LOCUS AP002957 87834 bp DNA PRI 26-APR-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:CTD-233717,
complete sequences.
ACCESSION AP002957
VERSION AP002957.2 GI:13810522
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:CTD-233717.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.
JOURNAL 2 (bases 1 to 87834)
DEFINITION Published Only in Database (2000) In press
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.
JOURNAL Direct Submission
TITLE Submitted (28-NOV-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suenhiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Apr 26, 2001 this sequence version replaced gi:11526584.
COMMENT Location/Qualifiers
FEATURES
Source
1..87834
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="CTD-233717"
/map="11q"

BASE COUNT 23650 a 20616 c 20127 g 23441 t
ORIGIN

Query Match 0.7%; Score 22; DB 9; Length 87834;
Best local Similarity 100.0%; Pred. No. 6.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1117 aaaaacaaacacacataaag 1138
|||||
Db 18515 AAAAAACAACAACATAAAG 18536

RESULT 19
AC020917 LOCUS AC020917 151088 bp DNA PRI 08-NOV-2000
DEFINITION Homo sapiens chromosome 19 clone CTD-2013N17, complete sequence.
ACCESSION AC020917
VERSION AC020917.4 GI:11120760
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 151088)
AUTHORS Doe Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Unpublished
AUTHORS Doe Joint Genome Institute.
JOURNAL 2 (bases 1 to 151088)
DEFINITION Published Only in Database (2000) In press
AUTHORS Doe Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 151088)
DEFINITION DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (08-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell

```

```

COMMENT
Drive, Walnut Creek, CA 94598, USA
On Nov 8, 2000 this sequence version replaced gi:7704958.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
STS Content:
SHGC-57513 G37313.
FEATURES
Source
1..151088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2013N17"

BASE COUNT 40571 a 38233 c 38536 g 33748 t
ORIGIN

Query Match 0.7%; Score 22; DB 9; Length 151088;
Best local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1778 attactattgactctcttga 1799
|||||
Db 117449 ATTACTATTGACTCTCTTGA 117470

RESULT 20
AC007951 LOCUS AC007951 standard; DNA; HTG; 151366 bp.
ACCESSION AC007951
VERSION AC007951.2
KEYWORDS HTG; HTGS-DRAFT; HTGS-PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 151366)
AUTHORS Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
Baker J., Baldwin J., Barna N., Beckerly R., Benn J., Brown A., Castle A.,
Cerny J., Colangelo M., Collins S., Collymore A., Cooke P., Dearellano K.,
DePatre E., Devon K., Dewar K., Donelan L., Doyle M., Ferreira P.,
Fitzhugh W., Forrest C., Funke K., Gage D., Galagan J., Gardyna S.,
Gilbert D., Grant G., Hagos B., Heaford A., Horton L., Howland J.C.,
Jones C., Kann L., Karatas A., Lehoczy J., Liu C., Locke K.,
Macdonald P., Marquis N., McEwan P., McGurk A., McKernan K., McLaughlin J.,
Meldrum J., Mollia M., Morris W., Morrow J., Mychaleckyj J., Naylor J.,
Niloff M., O'Connor T., O'Donnell P., Pavlin B., Peterson K., Pollara V.,
Riley R., Roberts D., Roy A., Severy P., Stange-Thomann N., Stojanovic N.,
Stone C., Subramanian A., Tsafaye S., Toriue-Miller I., Vassiliev H.,
Vo A., Wagner A., Wheeler J., Wu X., Wyman D., Ye W.J., Zody M.;
JOURNAL Submitted (02-JUL-1999) to the EMBL/GenBank/DBJ databases.
Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,

```

```

RL Cambridge, MA 02141, USA
XX
CC On Apr 1, 2000 this sequence version replaced gi:5332394.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: WIBR
CC Web site: http://www-seq.wi.mit.edu
CC Contact: sequence_submissions@genome.wi.mit.edu
CC ----- Project Information
CC Center project name: L878
CC Center clone name: 99.J.1
CC ----- Summary Statistics
CC Sequencing vector: M13; M77815; 96% of reads
CC Sequencing vector: Plasmid; n/a; %0.1% of reads
CC 3.80774032459426chemistry: Dye-primer-amerisham; 96% of reads
CC Chemistry: Dye-terminator Big Dye; 4% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 136641 bases at least Q40
CC Consensus quality: 145968 bases at least Q30
CC Consensus quality: 148942 bases at least Q20
CC Insert size: 15000; agarose-fp
CC Insert size: 150566; sum-of-connigs
CC Quality coverage:
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 9 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC 1 4869: contig of 4869 bp in length
CC 4870 4969: gap of 100 bp
CC 4970 10428: contig of 5459 bp in length
CC 10429 10528: gap of 100 bp
CC 10529 18565: contig of 8037 bp in length
CC 18566 18665: gap of 100 bp
CC 18666 26359: contig of 7594 bp in length
CC 26360 41183: contig of 14824 bp in length
CC 41184 41283: gap of 100 bp
CC 41284 53957: contig of 12674 bp in length
CC 53958 54057: gap of 100 bp
CC 54058 81728: contig of 27671 bp in length
CC 81729 81828: gap of 100 bp
CC 81829 114786: contig of 32958 bp in length
CC 114787 114886: gap of 100 bp
CC 114887 151366: contig of 36480 bp in length.
XX
FH Key
FH Location/Qualifiers
FH
FT source
FT 1. 151366
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /map="9"
FT /clone="RP11-99J1"
FT /clone_1id="RP11-99J1"
FT /clone_1id="RP11-99J1"
FT /note="assembly_fragment-clone_end:T7-vector_side:left"
FT 4970. 10428
FT /note="assembly_fragment"
FT 10529. 18565
FT /note="assembly_fragment"
FT 18666. 26359
FT /note="assembly_fragment-clone_end:SP6-vector_side:right"
FT 26360. 41183
FT /note="assembly_fragment"
FT 41284. 53957
FT /note="assembly_fragment"
FT misc_feature

```

```

FT misc_feature 54058. 81728
FT /note="assembly_fragment"
FT 81829. 114786
FT /note="assembly_fragment"
FT 11487. 151366
FT /note="assembly_fragment"
XX
SQ Sequence 151366 BP; 46140 A; 28393 C; 28257 G; 47776 T; 800 other;

Query Match 0.7%; Score 22; DB 33; Length 151366;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1060 ttccacttcacacacacaa 1081
Db 132511 TTCACCTCTCACACACAA 132532

RESULT 21
AC011980/c AC011980 153733 bp DNA HTG 12-MAR-2000
LOCUS Homo sapiens clone RP11-16H7, WORKING DRAFT SEQUENCE, 11 unordered
DEFINITION pieces.
ACCESSION AC011980 3 GI:7230122
VERSION AC011980
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 153733)
Britten,B., Linton,L., Nussbaum,C. and Lander,E.
Homo sapiens, clone RP11-16H7
Unpublished
2 (bases 1 to 153733)
Britten,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barne,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Funk,C., Heaford,A., Horton,L.,
Hollander,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Leloczky,J., Lieu,C., Locke,K., Macdonald,P., Marktas,N.,
McEwan,P., McGunk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,D., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Tirrelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6453966.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3543
Center clone name: 16_H_7
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 117952 bases at least Q40
Consensus quality: 136478 bases at least Q30
Consensus quality: 146705 bases at least Q20
Insert size: 157000; agarose-fp

```

Insert size: 152733; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      3262: contig of 3262 bp in length
*      3263: gap of 100 bp
*      3363 9412: contig of 6050 bp in length
*      9413 9512: gap of 100 bp
*      9513 14144: contig of 4632 bp in length
*      14145 14244: gap of 100 bp
*      14245 24193: contig of 9949 bp in length
*      24194 24293: gap of 100 bp
*      24294 37092: contig of 12799 bp in length
*      37093 37192: gap of 100 bp
*      37193 49508: contig of 12316 bp in length
*      49509 49608: gap of 100 bp
*      49609 60861: contig of 11253 bp in length
*      60862 60961: gap of 100 bp
*      60962 76206: contig of 15245 bp in length
*      76207 76306: gap of 100 bp
*      76307 89151: contig of 12845 bp in length
*      89152 89251: gap of 100 bp
*      89252 107693: contig of 18442 bp in length
*      107694 107793: gap of 100 bp
*      107794 153733: contig of 45940 bp in length.

```

FEATURES

```

source
1. 153733
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone_lib="RP11-16H7"
   /clone_1lb="RP11-11 Human Male BAC"
   1. 3262
      /note="assembly_fragment"
misc_feature
3363. 9412
   /note="assembly_fragment"
misc_feature
9513. 14144
   /note="assembly_fragment"
misc_feature
14245. 24193
   /note="assembly_fragment"
misc_feature
24294. 37092
   /note="assembly_fragment"
misc_feature
37193. 49508
   /note="assembly_fragment"
misc_feature
49609. 60861
   /note="assembly_fragment"
misc_feature
clone_end:T7
   /note="assembly_fragment"
vector_side:right"
60962. 76206
   /note="assembly_fragment"
misc_feature
76307. 89151
   /note="assembly_fragment"
misc_feature
clone_end:SP6
   /note="assembly_fragment"
vector_side:right"
89252. 107693
   /note="assembly_fragment"
misc_feature
107794. 153733
   /note="assembly_fragment"
BASE COUNT 42218 a 34152 c 33610 g 42710 t 1043 others
ORIGIN

```

Query Match 0.7%: Score 22; DB 2; Length 153733;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1117 aaaaacaacacacataaag 1138
Db 120518 AAAAACAACACAACATAAG 120497

```

RESULT 22
AL592213/c
LOCUS
DEFINITION
Homo sapiens chromosome 9 clone RP11-99J1, ** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
AL592213 AC007951
VERSION
AL592213.6 GI:15131900.
KEYWORDS
HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 155531)
REFERENCE
1 Hammond, S.
Direct Submission
Submitted (03-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Aug 9, 2001 this sequence version replaced gi:15029558.
Draft Sequence Produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba89J1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 30% of reads
Sequencing vector: plasmid; 108752; 69% of reads
Chemistry: Dye-terminator ET-amersham; 3% of reads
Dye-terminator Big Dye; 68% of reads
Chemistry: Dye-primer-amersham; 28% of reads
Consensus quality: 155436 bases at least Q40
Consensus quality: 155453 bases at least Q30
Consensus quality: 155482 bases at least Q20
Insert size: 155531; sum-of-contigs
Insert size: 144681; 16.2% error; agarose-fp
Quality coverage: 15.49% in Q20 bases; sum-of-contigs Quality
coverage: 16.65% in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 155531
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="9"
   /clone_lib="RP11-99J1"
   /clone_1lb="RP11-11.1"
   1. 155531
      /note="assembly_fragment:03518
clone_end:SP6
   /note="assembly_fragment"
vector_side:right"
misc_feature
vector_side:right"
BASE COUNT 50143 a 29768 c 28721 g 46899 t
ORIGIN

```

Query Match 0.7%: Score 22; DB 2; Length 155531;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1060 ttccactctcacacacaacaa 1081
|||||
Db 38569 TTCACCTCTCACACACAACAA 38548

RESULT 23
AP002958/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-16H7 map 11q, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION AP002958
VERSION AP002958.1 GI:11526585
KEYWORDS HTG; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-16H7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1. (bases 1 to 156230)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.
Homo sapiens 156,230 genomic DNA of 11q
2 (bases 1 to 156230)
Published Only in Database (2000) In press
TITLE
JOURNAL Submitted (28-NOV-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
AUTHORS
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.
DIRECT SUBMISSION
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
Project Information
Center project name: Rpl1-16H7
Center clone name: Rpl1-16H7
Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 152390 bases at least Q40
Consensus quality: 153962 bases at least Q30
Consensus quality: 154642 bases at least Q20
Insert size: 154830; sum-of-contigs
Quality coverage: 8.02x in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently consists of
15 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.

1 32099 contig of 32099 bp in length
1 32200 46804 contig of 14605 bp in length
62728 contig of 15824 bp in length
62829 76891 contig of 14063 bp in length
76992 90383 contig of 13392 bp in length
90484 100561 contig of 10078 bp in length
100662 114890 contig of 14229 bp in length
114991 125640 contig of 10650 bp in length
125741 134470 contig of 8730 bp in length
134571 140147 contig of 5577 bp in length
140248 147375 contig of 7128 bp in length
147476 150473 contig of 2998 bp in length
150574 153025 contig of 2452 bp in length
153126 153126 155065 contig of 1940 bp in length
155166 156230 contig of 1065 bp in length
155166

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces

is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 32099: contig of 32099 bp in length
* 32100 32199: gap of 100 bp
* 32200 46804: contig of 14605 bp in length
* 46805 46904: gap of 100 bp
* 46905 62728: contig of 15824 bp in length
* 62729 62828: gap of 100 bp
* 62829 76891: contig of 14063 bp in length
* 76892 76991: gap of 100 bp
* 76992 90383: contig of 13392 bp in length
* 90384 90483: gap of 100 bp
* 90484 100561: contig of 10078 bp in length
* 100562 100661: gap of 100 bp
* 100662 114890: contig of 14229 bp in length
* 114891 114990: gap of 100 bp
* 114991 125640: contig of 10650 bp in length
* 125641 125740: gap of 100 bp
* 125741 134470: contig of 8730 bp in length
* 134471 134570: gap of 100 bp
* 134571 140147: contig of 5577 bp in length
* 140148 140247: gap of 100 bp
* 140248 147375: contig of 7128 bp in length
* 147376 147475: gap of 100 bp
* 147476 150473: contig of 2998 bp in length
* 150474 150573: gap of 100 bp
* 150574 153025: contig of 2452 bp in length
* 153026 153125: gap of 100 bp
* 153126 155065: contig of 1940 bp in length
* 155066 155165: gap of 100 bp
* 155166 156230: contig of 1065 bp in length.
Location/Qualifiers
1. 156230
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-16H7"
1. 32099
/note="assembly-fragment"
32200. 46804
/note="assembly-fragment clone_end:SP6 vector_side:left"
46905. 62728
/note="assembly-fragment"
62829. 76891
/note="assembly-fragment"
76992. 90383
/note="assembly-fragment"
90484. 100561
/note="assembly-fragment clone_end:r7 vector_side:left"
100662. 114890
/note="assembly-fragment"
114991. 125640
/note="assembly-fragment"
125741. 134470
/note="assembly-fragment"
134571. 140147
/note="assembly-fragment"
140248. 147375
/note="assembly-fragment"
147476. 150473
/note="assembly-fragment"
150574. 153025
/note="assembly-fragment"
153126. 155065
/note="assembly-fragment"
155166. 156230
/note="assembly-fragment"
155166

BASE COUNT 43180 a 34058 c 34268 g 43324 t 1400 others

ORIGIN

Query Match 0.7%; Score 22; DB 2; Length 156230;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1117 aaaaacacacacacataaag 1138
AAAAAAAAAAAAAAAAAAAAA
Db 73708 AAAAAACAAACAAACATAAAG 73687

RESULT 24
AC079194/c DNA HTG 22-NOV-2000
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-475C4 map 11, WORKING DRAFT
AC079194
AC079194.2 GI:11276162
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 159946)
Birren,B., Linton,L., Nussbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-475C4
Unpublished
2 (bases 1 to 159946)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,Y., Beda,F., Boguslavsky,L.,
Boukhalil,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
Choelel,Y., Collangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Flitzhug,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Larocque,K.,
Lamatares,R., Landers,T., Lenocksky,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Souarez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (23-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2000 this sequence version replaced gi:9886030.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10601
Center clone name: 475_C-4
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 139893 bases at least Q40
Consensus quality: 150490 bases at least Q30
Consensus quality: 154263 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 156346; sum-of-ctrls

Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
2598 2697: gap of 100 bp
2698 3739: contig of 1042 bp in length
3740 3839: gap of 100 bp
3840 4797: contig of 958 bp in length
4798 4897: gap of 100 bp
4898 6197: contig of 1300 bp in length
6198 6297: gap of 100 bp
6298 8221: contig of 1924 bp in length
8222 8321: gap of 100 bp
8322 9558: contig of 1237 bp in length
9559 9658: gap of 100 bp
9659 11181: contig of 1523 bp in length
11182 12622: contig of 1341 bp in length
12623 12722: gap of 100 bp
12723 14374: contig of 1652 bp in length
14375 14474: gap of 100 bp
14475 37771: contig of 23297 bp in length
37772 37871: gap of 100 bp
37872 39354: contig of 1483 bp in length
39355 39454: gap of 100 bp
39455 40854: contig of 1400 bp in length
40855 40954: gap of 100 bp
40955 42709: contig of 1755 bp in length
42710 42809: gap of 100 bp
42810 44167: contig of 1358 bp in length
44168 44267: gap of 100 bp
44268 46404: contig of 2137 bp in length
46405 46504: gap of 100 bp
46505 48310: contig of 1806 bp in length
48311 48410: gap of 100 bp
48411 51136: contig of 2726 bp in length
51137 51236: gap of 100 bp
51237 53831: contig of 2595 bp in length
53832 53931: gap of 100 bp
53932 56439: contig of 2508 bp in length
56440 56539: gap of 100 bp
56540 58998: contig of 2459 bp in length
58999 59098: gap of 100 bp
59099 63491: contig of 4393 bp in length
63492 63591: gap of 100 bp
63592 66743: contig of 3152 bp in length
66744 66843: gap of 100 bp
66844 69817: contig of 2974 bp in length
69818 69917: gap of 100 bp
69918 73103: contig of 3186 bp in length
73104 73203: gap of 100 bp
73204 77583: contig of 4380 bp in length
77584 77683: gap of 100 bp
77684 83310: contig of 5627 bp in length
83311 83410: gap of 100 bp
83411 87029: contig of 3619 bp in length
87030 87129: gap of 100 bp
87130 92085: contig of 4956 bp in length
92086 92185: gap of 100 bp
92186 97647: contig of 5462 bp in length
97648 97747: gap of 100 bp
97748 101467: contig of 3720 bp in length
101468 101567: gap of 100 bp
101568 106778: contig of 5211 bp in length
106779 106878: gap of 100 bp

```

* 106879 114040: contig of 7162 bp in length
* 114041 114140: gap of 100 bp
* 114141 123890: contig of 9750 bp in length
* 123891 123990: gap of 100 bp
* 123991 130670: contig of 6680 bp in length
* 130671 130770: gap of 100 bp
* 130771 139820: contig of 9050 bp in length
* 139821 139920: gap of 100 bp
* 139921 158959: contig of 19039 bp in length
* 158960 159059: gap of 100 bp
* 159060 159946: contig of 887 bp in length.

```

```

FEATURES
source
1..159946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-475C4"
/clone_1lb="RPC1-11 Human Male BAC"
1..2597

```

```

misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left"
2698..3739
/misc_feature /note="assembly_fragment"
3840..4797
/misc_feature /note="assembly_fragment"
4898..6197
/misc_feature /note="assembly_fragment"
6298..8221
/misc_feature /note="assembly_fragment"
8322..9558
/misc_feature /note="assembly_fragment"
9659..11181
/misc_feature /note="assembly_fragment"
11282..12622
/misc_feature /note="assembly_fragment"
12723..14374
/misc_feature /note="assembly_fragment"
14475..37771
/misc_feature /note="assembly_fragment"
37872..39354
/misc_feature /note="assembly_fragment"
39455..40854
/misc_feature /note="assembly_fragment"
40955..42709
/misc_feature /note="assembly_fragment"
42810..44167
/misc_feature /note="assembly_fragment"
44268..46404
/misc_feature /note="assembly_fragment"
46505..48310
/misc_feature /note="assembly_fragment"
48411..51136
/misc_feature /note="assembly_fragment"
51237..53831
/misc_feature /note="assembly_fragment"
53932..56439
/misc_feature /note="assembly_fragment"
56540..58998
/misc_feature /note="assembly_fragment"
59099..63491
/misc_feature /note="assembly_fragment"
63592..66743
/misc_feature /note="assembly_fragment"
66844..69817
/misc_feature /note="assembly_fragment"
69918..73103

```

```

Query Match 0.7%; Score 22; DB 2; Length 159946;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1117 aaaaacaaacacataaag 1138
|||||
Db 150590 AAAAACAAACACAAACATAAG 150569

```

```

RESULT 25
AL161913/c
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-64P11, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.

```

```

ACCESSION AL161913
VERSION AL161913.10 GI:14329900
KEYWORDS HTGS: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186413)
TITLE Mashreghi-Mohammadi, M.
JOURNAL Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14141268.
----- Genome Center

```

COMMENT

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA64P11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing Vector: plasmid; L08752: 100% of reads
Chemistry: Dye-terminator ET-amersham; 7% of reads
Dye-terminator Big Dye; 92% of reads
Consensus quality: 185600 bases at least Q40
Consensus quality: 185986 bases at least Q30
Consensus quality: 186182 bases at least Q20
Insert size: 186313; sum-of-contigs
Insert size: 140000; 37.9% error; agarose-fp
Quality coverage: 9.01x in Q20 bases; sum-of-contigs Quality
coverage: 12.46x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 168456: contig of 168456 bp in length
* 168457 168556: gap of 100 bp
* 168557 186413: contig of 17857 bp in length.

```

FEATURES

```

source
1..186413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-64P11"
/clone_1lb="RPC1-11.1"
1..168456
/note="assembly_fragment:00638
clone_end:SP6
vector_side:left"
168557..186413
/note="assembly_fragment:00456"
misc_feature /note="assembly_fragment:00456"
100 others
BASE COUNT 60691 a 35149 c 33874 g 56599 t
ORIGIN

```

```

Query Match 0.7%; Score 22; DB 2; Length 186413;

```

Best Local Similarity 100.0%; Pred No. 6.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1060 ttccactctcacacacaacaa 1081
|||||
Db 174706 ttccactctcacacacaacaa 174685

RESULT 26
AL592438 194533 bp DNA HTG 02-AUG-2001
LOCUS Homo sapiens chromosome 9 clone RP11-18766, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL592438
VERSION AL592438.6 GI:15021052
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 194533)
JOURNAL Direct Submission
Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14787678.
KEYWORDS Genom Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA18766

COMMENT
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 9% of reads
Sequencing vector: plasmid; L08752; 90% of reads
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer-amerisham; 9% of reads
Consensus quality: 194192 bases at least Q40
Consensus quality: 194424 bases at least Q30
Consensus quality: 194498 bases at least Q20
Insert size: 194533; sum-of-contents
Insert size: 166314; agarose-fp
Quality coverage: 9.01x in Q20 bases; sum-of-contents Quality
coverage: 10.54x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. 194533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-18766"
/clone_lib="RPCT-11.1"
misc_feature
1. 194533
/note="assembly_fragment:02981"
BASE COUNT 61649 a 36707 c 36581 g 59596 t
ORIGIN

Query Match .0.7%; Score 22; DB 2; Length 194533;
Best Local Similarity 100.0%; Pred No. 6.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1060 ttccactctcacacacaacaa 1081
|||||
Db 194370 ttccactctcacacacaacaa 194349

RESULT 27
BTCSK35 7595 bp DNA MAM 24-JUN-2000
LOCUS Bovine gene for kappa-casein exons 3-5.
DEFINITION X14908 X14326
ACCESSION X14908.1 GI:180
VERSION X14908.1 GI:180
KEYWORDS casein; kappa-casein; repetitive sequence: Alu-like repetitive
sequence; restriction fragment linked polymorphism.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 7595)
Alexander, L.J., Stewart, A.F., Mackinlay, A.G., Kapelinskaya, T.V.,
Tkach, T.M. and Gorodetsky, S.I.
Isolation and characterization of the bovine kappa-casein gene
Eur. J. Biochem. 178 (2), 395-401 (1988)
2 (bases 1 to 7595)
Vaiman, D., Mercier, D., Mozamli-Goudarzi, K., Eggen, A.,
Ciampolini, R., Leplingle, A., Veimola, R., Kaukinen, J., Varvio, S.L.,
Martin, P. et al.
A set of 99 cattle microsatellites: characterization, synteny
mapping, and polymorphism
Mamm. Genome 5 (5), 288-297 (1994)
94355772
3 (bases 1 to 7595)
Mackinlay, A.G.
Direct Submission
Submitted (16-MAR-1989) A.G. Mackinlay, The University of New South
Wales, P.O. Box 1 Kensington New South Wales Australia
Location/Qualifiers
1. 7595
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="EMBL3, EMBL4."
2886. 2891
/label="sigstop"
/usedin=X14907.kc_slg
2886. 2918
/label="ex3"
/usedin=X14907.kc_cds
/usedin=X14906.kc_mrna
join(2892. 2918, 4930. 5412)
/label="kc_mat"
/product="kappa-casein"
4930. 5412
/label="stop"
/usedin=X14907.kc_cds
4930. 5446
/label="ex4"
/usedin=X14906.kc_mrna
5309
/note="c is t in kappa-casein B variant; changes acc (Thr)
to atc (Tle)"
5345
/note="a is c in kappa-casein B variant; changes gat (Asp)
to gct (Ala)"
5406
/note="a is g in kappa-casein B variant; loss of PstI
site"
5413
/note="a is t in kappa-casein B variant"
7296. 7416
/label="ex5"
/usedin=X14906.kc_mrna
7382
/note="t is c in kappa-casein B variant; loss of BglII
site, creates MspI site"
7416
7484

FEATURES
source
1. 7595
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="EMBL3, EMBL4."
2886. 2891
/label="sigstop"
/usedin=X14907.kc_slg
2886. 2918
/label="ex3"
/usedin=X14907.kc_cds
/usedin=X14906.kc_mrna
join(2892. 2918, 4930. 5412)
/label="kc_mat"
/product="kappa-casein"
4930. 5412
/label="stop"
/usedin=X14907.kc_cds
4930. 5446
/label="ex4"
/usedin=X14906.kc_mrna
5309
/note="c is t in kappa-casein B variant; changes acc (Thr)
to atc (Tle)"
5345
/note="a is c in kappa-casein B variant; changes gat (Asp)
to gct (Ala)"
5406
/note="a is g in kappa-casein B variant; loss of PstI
site"
5413
/note="a is t in kappa-casein B variant"
7296. 7416
/label="ex5"
/usedin=X14906.kc_mrna
7382
/note="t is c in kappa-casein B variant; loss of BglII
site, creates MspI site"
7416
7484

misc_feature
4930. 5412
/label="stop"
/usedin=X14907.kc_cds
4930. 5446
/label="ex4"
/usedin=X14906.kc_mrna
5309
/note="c is t in kappa-casein B variant; changes acc (Thr)
to atc (Tle)"
5345
/note="a is c in kappa-casein B variant; changes gat (Asp)
to gct (Ala)"
5406
/note="a is g in kappa-casein B variant; loss of PstI
site"
5413
/note="a is t in kappa-casein B variant"
7296. 7416
/label="ex5"
/usedin=X14906.kc_mrna
7382
/note="t is c in kappa-casein B variant; loss of BglII
site, creates MspI site"
7416
7484


```
/note="t is c in variant clone"
variation 7486
/note="t is c in variant clone"
variation 7504
/note="t is c in variant clone"
variation 7554
/note="g is a in variant clone"
variation 7570..7571
/note="c is inserted in variant clone"
BASE COUNT 2646 a 1292 c 1182 g 2475 t
ORIGIN

Query Match 0.7%; Score 21; DB 4; Length 7595;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 380 ttcttacttgaaagcact 400
|||||
Db 6007 TTTTAACTTGAAAGCACT 6027

RESULT 28
AC005199 37906 bp DNA PRI 30-JUN-1998
LOCUS Homo sapiens chromosome 16, cosmid clone 330D11 (LANL), complete
DEFINITION sequence.
ACCESSION AC005199
VERSION AC005199.1 GI:3273387
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 37906)
REFERENCE 1
AUTHORS Rieke,D.O., Bruce,D., Muntt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnicke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Mistra,M. and Deaven,L.
Sequencing of Human Chromosome 16p13.3
Unpublished
2 (bases 1 to 37906)
REFERENCE 2
AUTHORS Rieke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 37906)
AUTHORS Rieke,D.O., Bruce,D., Muntt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnicke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Mistra,M. and Deaven,L.
Direct Submission
TITLE Submitted (30-JUN-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
JOURNAL
REFERENCE
AUTHORS
FEATURES
Source Location/Qualifiers
1..37906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="330D11"
/chromosome="16"
/map="16p13.3"
2548..2792
/repeat_region="MLT1"
4130..4301
/rpt_family="Alu"
5671..5997
/note="97% identity EST yb12f02.r1"
/db_xref="dbEST:747043"
6373..6501
/repeat_region="MER25"
/repeat_region="MER25"
```

```
repeat_region 6572..6858
/rpt_family="Alu"
repeat_region 7074..7299
/rpt_family="L1"
repeat_region 7509..7828
/rpt_family="L1"
repeat_region 7898..10788
/rpt_family="L1"
repeat_region 11765..11854
/rpt_family="MER21"
repeat_region complement(12556..12682)
/rpt_family="MIR2"
repeat_region complement(13150..13299)
/rpt_family="L1"
repeat_region 13754..14096
/rpt_family="Alu"
repeat_region 14119..14293
/rpt_family="MER5"
repeat_region complement(14450..14651)
/rpt_family="MIR"
repeat_region 15660..15985
/rpt_family="Alu"
repeat_region complement(17523..17908)
/rpt_family="MSTa"
repeat_region 18375..18547
/rpt_family="MIR"
misc_feature complement(18614..18884)
/note="Grail 2 excellent exon, frame 0"
repeat_region 19269..19921
/rpt_family="MER44C"
20666..27120
/rpt_family="L1"
repeat_region 22343..22577
/rpt_family="MER25"
30116..30259
/rpt_family="HSAT1"
repeat_region complement(30385..30580)
/rpt_family="HSAT1"
repeat_region complement(30678..30786)
/rpt_family="HSAT1"
repeat_region complement(31921..32340)
/rpt_family="MLT1"
33481..33615
/rpt_family="Alu"
repeat_region 33613..34213
/rpt_family="L1"
repeat_region 34664..34710
/rpt_family="MIR"
misc_feature complement(34811..35200)
/note="99% identity A0009624 BAC end sequence"
35732..36036
/rpt_family="Alu"
repeat_region 36016..36036
/note="(A)21"
/rpt_type=tandem
repeat_region complement(36219..36374)
/rpt_family="MLT1d"
BASE COUNT 13076 a 7620 c 7679 g 9531 t
ORIGIN

Query Match 0.7%; Score 21; DB 9; Length 37906;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1977 cagagctggagctacact 1997
|||||
Db 1512 CAGGACTGTGAGGTCACACT 1532

RESULT 29
AC017948/c
```

```

LOCUS       AC017948      45496 bp      DNA           09-DEC-1999
DEFINITION   Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***. In ordered
ACCESSION   AC017948
VERSION     AC017948.1 GI:6553242
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 45496)
AUTHORS     Adams,M. and Venter,J.C.
TITLE       Direct Submission
JOURNAL     Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212758 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.

FEATURES
  source
    1..45496
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"

BASE COUNT   13595 a      9140 c      9064 g      13697 t

ORIGIN
Query Match      0.7%: Score 21; DB 2; Length 45496;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1631 aaattctaaagccataaaca 1651
Db 29997 AAATTTGAAAGCCATAACA 29977

RESULT 30
AC027055/c
LOCUS       AC027055      76526 bp      DNA           13-JUL-2000
DEFINITION   Homo sapiens chromosome 3 clone RP11-334K8 map 3, LOW-PASS SEQUENCE
ACCESSION   AC027055
VERSION     AC027055.1 GI:7329416
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 76526)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-334K8
Unpublished
2 (bases 1 to 76526)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArliano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,L.,
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Margus,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Melchum,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

```

```

TITLE
JOURNAL
COMMENT
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L7376
Center clone name: 334_K_8
-----
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overall relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1
804 903: contig of 803 bp in length
804 903: gap of 100 bp
904 1690: contig of 787 bp in length
1691 1790: gap of 100 bp
1791 2559: contig of 769 bp in length
2560 2659: gap of 100 bp
2660 3445: contig of 786 bp in length
3446 3545: gap of 100 bp
3546 4348: contig of 803 bp in length
4349 4448: gap of 100 bp
4449 5243: contig of 795 bp in length
5244 5343: gap of 100 bp
5344 6119: contig of 776 bp in length
6120 6219: gap of 100 bp
6220 6998: contig of 779 bp in length
6999 7098: gap of 100 bp
7099 7884: contig of 786 bp in length
7885 7984: gap of 100 bp
7985 8766: contig of 782 bp in length
8767 8866: gap of 100 bp
8867 9669: contig of 803 bp in length
9670 9769: gap of 100 bp
9770 10559: contig of 790 bp in length
10560 10659: gap of 100 bp
10660 11449: contig of 790 bp in length
11450 11549: gap of 100 bp
11550 12330: contig of 781 bp in length
12331 12430: gap of 100 bp
12431 13222: contig of 792 bp in length
13223 13322: gap of 100 bp
13323 14098: contig of 776 bp in length
14099 14198: gap of 100 bp
14199 14992: contig of 794 bp in length
14993 15092: gap of 100 bp
15093 15881: contig of 789 bp in length
15882 15981: gap of 100 bp
15982 16771: contig of 790 bp in length
16772 16871: gap of 100 bp
16872 17655: contig of 784 bp in length
17656 17755: gap of 100 bp
17756 18549: contig of 794 bp in length
18550 18649: gap of 100 bp
18650 19440: contig of 791 bp in length

```

```

* 19441 19540: gap of 100 bp
* 19541 20335: contig of 795 bp in length
* 20336 20435: gap of 100 bp
* 20436 21236: contig of 801 bp in length
* 21237 21336: gap of 100 bp
* 21337 22140: contig of 804 bp in length
* 22141 22240: gap of 100 bp
* 22241 23041: contig of 801 bp in length
* 23042 23141: gap of 100 bp
* 23142 23942: contig of 801 bp in length
* 23943 24042: gap of 100 bp
* 24043 24840: contig of 798 bp in length
* 24841 24940: gap of 100 bp
* 24941 25729: contig of 789 bp in length
* 25730 25829: gap of 100 bp
* 25830 26603: contig of 774 bp in length
* 26604 26703: gap of 100 bp
* 26704 27480: contig of 777 bp in length
* 27481 27580: gap of 100 bp
* 27581 28381: contig of 801 bp in length
* 28382 28481: gap of 100 bp
* 28482 29279: contig of 798 bp in length
* 29280 29379: gap of 100 bp
* 29380 30167: contig of 788 bp in length
* 30168 30267: gap of 100 bp
* 30268 31056: contig of 789 bp in length
* 31057 31156: gap of 100 bp
* 31157 31955: contig of 799 bp in length
* 31956 32055: gap of 100 bp
* 32056 32844: contig of 789 bp in length
* 32845 32944: gap of 100 bp
* 32945 33731: contig of 787 bp in length
* 33732 33831: gap of 100 bp
* 33832 34582: contig of 751 bp in length
* 34583 34682: gap of 100 bp
* 34683 35480: contig of 798 bp in length
* 35481 35580: gap of 100 bp
* 35581 36381: contig of 801 bp in length
* 36382 36481: gap of 100 bp
* 36482 37276: contig of 795 bp in length
* 37277 37376: gap of 100 bp
* 37377 38182: contig of 806 bp in length
* 38183 38282: gap of 100 bp
* 38283 39072: contig of 790 bp in length
* 39073 39172: gap of 100 bp
* 39173 39983: contig of 811 bp in length
* 39984 40083: gap of 100 bp
* 40084 40873: contig of 790 bp in length
* 40874 40973: gap of 100 bp
* 40974 41759: contig of 786 bp in length
* 41760 41859: gap of 100 bp
* 41860 42652: contig of 793 bp in length
* 42653 42752: gap of 100 bp
* 42753 43560: contig of 808 bp in length
* 43561 43660: gap of 100 bp
* 43661 44448: contig of 788 bp in length
* 44449 44548: gap of 100 bp
* 44549 45335: contig of 787 bp in length
* 45336 45435: gap of 100 bp
* 45436 46220: contig of 785 bp in length
* 46221 46320: gap of 100 bp
* 46321 47108: contig of 788 bp in length
* 47109 47208: gap of 100 bp
* 47209 47983: contig of 775 bp in length
* 47984 48083: gap of 100 bp
* 48084 48863: contig of 780 bp in length
* 48864 48963: gap of 100 bp
* 48964 49755: contig of 792 bp in length
* 49756 49855: gap of 100 bp
* 49856 50659: contig of 804 bp in length
* 50660 50759: gap of 100 bp
* 50760 51542: contig of 783 bp in length
* 51543 51642: gap of 100 bp

```

```

* 51643 52427: contig of 785 bp in length
* 52428 52527: gap of 100 bp
* 52528 53328: contig of 801 bp in length
* 53329 53428: gap of 100 bp
* 53429 54211: contig of 783 bp in length
* 54212 54311: gap of 100 bp
* 54312 55088: contig of 777 bp in length
* 55089 55188: gap of 100 bp
* 55189 55959: contig of 771 bp in length
* 55960 56059: gap of 100 bp
* 56060 56859: contig of 800 bp in length
* 56860 56959: gap of 100 bp
* 56960 57762: contig of 803 bp in length
* 57763 57862: gap of 100 bp
* 57863 58648: contig of 786 bp in length
* 58649 58748: gap of 100 bp
* 58749 59547: contig of 799 bp in length
* 59548 59647: gap of 100 bp
* 59648 60435: contig of 788 bp in length
* 60436 60535: gap of 100 bp
* 60536 61333: contig of 798 bp in length
* 61334 61433: gap of 100 bp

```

```

Query Match          0.7% Score 21: DB 2: Length 76526;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1111 atacgagaaacacacacacaa 1131
Db 4183 ATACGAGAAACACACACACAA 4163

```

RESULT 31

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

FEATURES

source

```

1. 80346
/organism="Arabidopsis thaliana"
/Variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
/complement(join(2..154,234..572))
/gene="M4E13.10"
2..572
/complement(join(<2..154,234..572))
/gene="M4E13.10"
/notice="Similarity to Arabidopsis thaliana AT.I.24-7,
PATCH:G1532169"
/codon_start=1
/product="EF-1 alpha-like protein (fragment)"

```

/protein_id="CAAI1760.1"
 /db_xref="GI:445135"
 /translation="MGFTSSSSSSALKMLGFYAVWVQSIGNNYTPSNYSALK
 SLNNIQLNELNSLVAKDVGAKFGLAGLASDRLSPVILLISPEGLGAGYQMWLYV
 SRTIPIPYMQKVFILCMGNSITWMNTAVLYCIRNRRNRGPGVSLIKGYGLSTIA
 ITFD"
 complement(2..154)
 /gene="MAE13.10"
 /number=1
 complement(155..233)
 /number=1
 complement(234..572)
 /gene="MAE13.10"
 /number=2
 join(4678..4776,5251..5332,5481..5528,5618..5824,
 5954..6029,6297..6459)
 /gene="MAE13.20"
 /note="strong similarity to cyclophilin, Arabidopsis
 thaliana, P4TCHX:G244375
 Contains Cyclophilin-type peptidyl-prolyl cis-trans
 isomerase signature & profile, [YKGPFRHISGFVIQGG]"
 /codon_start=1
 /product="cyclophilin-like protein"
 /protein_id="CAAI1761.1"
 /db_xref="GI:2924507"
 /db_xref="SPTREMBL:049605"
 /translation="MRREISFLQPRCLLLVALLTFLVFALEFNTGKDEEKQVIEDHE
 ITRRVFLVDIDGRIYGLGYVPTVFNFRALCTGKGTSSGKPLHYKTP
 FHRISGFVIQGGDIHGCKSSDSITGCFNDENKIOGSHAGVAMANTGDSNGS
 QFRTTYKASMLGEHNVLCVIQGMNFAIGAGTTSKFRKRVLIADSEITRD
 KWDEER
 4678..4776
 /gene="MAE13.20"
 /number=1
 4678..6459
 /gene="MAE13.20"
 4777..5250
 /gene="MAE13.20"
 /number=1
 5251..5332
 /gene="MAE13.20"
 /number=2
 5333..5460
 /gene="MAE13.20"
 /number=2
 5481..5528
 /gene="MAE13.20"
 /number=3
 5529..5617
 /gene="MAE13.20"
 /number=3
 5618..5824
 /gene="MAE13.20"
 /number=4
 5825..5953
 /gene="MAE13.20"
 /number=4
 5954..6029
 /gene="MAE13.20"
 /number=5
 6030..6296
 /gene="MAE13.20"
 /number=5
 6297..6459
 /gene="MAE13.20"
 /number=6
 9714..9955
 /gene="MAE13.30"
 /number=1
 9714..10406
 /gene="MAE13.30"
 join(9714..9955,10256..10406)

/gene="MAE13.30"
 /note="strong similarity to actin depolymerizing factor1,
 Arabidopsis thaliana, P4TCHX:G1408471
 Contains Actin-depolymerizing proteins signature,
 [P6ASIRREKMYATKSLRL]"
 /codon_start=1
 /product="actin depolymerizing factor-like protein"
 /protein_id="CAAI1762.1"
 /db_xref="GI:2924508"
 /db_xref="SPTREMBL:049606"
 /translation="MTDDCKKSPMEKKKRVHRYVYKLEESRRKTVDKVAGESY
 DDLAASLPDDDCRYAFEDPDYTVNDCRSKSLFFITWSEASRIEKKMYATKSGLR
 RVLGVHYEDQATPDTEMGFKI0DRAK"
 9956..10255
 /gene="MAE13.30"
 /number=1
 10256..10406
 /gene="MAE13.30"
 /number=2
 11099..11172
 /gene="MAE13.40"
 12742..15036
 /gene="MAE13.40"
 complement(12742..15036)
 /gene="MAE13.40"
 /number=1
 complement(12742..15036)
 /gene="MAE13.40"
 complement(12742..15036)
 /gene="MAE13.40"
 /note="strong similarity to subtilisin-like proteinase,
 Arabidopsis thaliana, PIR2:S52770
 Contains Serine proteases, F14303, F14305"
 [GSMACPHVSG]
 contains EST gb:F14303, F14305"
 /codon_start=1
 /product="subtilisin proteinase-like"
 /protein_id="CAAI1763.1"
 /db_xref="GI:2924509"
 /db_xref="SPTREMBL:049607"
 /translation="MASTIVLLFLSFPFISFAASQAQKTFIFRIDGSMPSIFPTH
 YHWSTEFAPESRIYHVHTVFHGFSAVTPEDADNLRNHPAVLAFEDRRELEHTTR
 SPOFLGIONKGIKMSSESDYGIICGPTGTGIMPRRSPDNLGPIPRMRGCESG
 ARFSFRCNKRKIIGARFRFAKGGQAAVIGIKNTVEPLSPRDGNGHTTSIACGSHA
 FKASMSGVAGVAKGVAARKARIAAKVCKKDSGCDSDILAAFDAAVRGVVISISI
 GGGDJTISPYLDPLAIGSYGAASKGIEFVSSAGNDEGPKGMSVTMLAPWTVVAGSTI
 DRRFPADAILGCGHRLRGVSLAGVPLNRPMPVYVPGKSGMSALCMENTLDPROV
 RCKIVICDRGSSPRVAKGLVKKAGVGWMLNGASNGEGLVGDHLLIPACAVSNEG
 DRKATAYASHPAPIASIDFRGTIVGIRKRPVYIASRSGRPNLSPEILKPOLIARGVN
 ILAAMTDAVGPTGLSPDPKRTENILISGTSMACPHVSGAALLKSHHPWSPAVIRSA
 MMTTNLVNLSNRSLLIDESGKSATPYDGSGLNLNGRMNGLVYDITNDYITFLC
 SIGYGKTIQVITRPEVRCPTTRKPSGMLNPSITAVPPTNRGLVSKTVLRATANV
 GOAEAVYRARIESPRGVTVKPPRLVFTSAVKRRSYAVTVNTRNVVLGETGAVFG
 SYTWEPDGKHHVRSPIVTVQMDTL"
 complement(17183..17744)
 /gene="MAE13.50"
 /number=1
 17183..18102
 /gene="MAE13.50"
 complement(join(17183..17744,17840..18102))
 /gene="MAE13.50"
 complement(join(17183..17744,17840..18102))
 /gene="MAE13.50"
 /note="strong similarity to myb-related protein Y49,
 Arabidopsis thaliana, PIR1:S58292
 Contains Myb DNA-binding domain repeat signatures,
 [WTKEDDKL][WSLIATRLRGTRDREIKYNNHY], Aldo/keto reductase
 family signatures, [LPSSAGLQCGSSCR], AtP/GTP-binding
 site motif A (P-loop), [GGRGKGS]"
 /codon_start=1
 /product="MYB-like protein"
 /protein_id="CAAI1764.1"
 /db_xref="GI:2924510"
 /db_xref="SPTREMBL:049608"

/translation="MGRSPCEKDHNTKNGAWTKFEEDKLISYKIHAGEGCWRSLPRSA
GLQRGKSCRLRWLNLYRDPDKRGNFTLEEDDLIKLHSLGKNWLSLARTLPQRTN
EIKYWNTHVYKRLKLRNGIDPATRPINETSTODSSSSSKTEEDPLVKIISFGPOLK
IANFGDERIOKRVKSVYAEKCLDLNLELRISPPWOKLHDERLNRGRYKYSAR
FGFNGKDECSNNKVCOTEDSSSSYSTDISSISGIDFGLNFRVLDPSTLEMK-
complement(17745..17839)
/number=1
exon complement(17840..18102)
/gene="M4E13.50"
/number=2
gene complement(join(20820..21014,21397..21499,21644..21723,
21807..21889,22003..22051,22140..22205,22752..22801,
22966..23090,23242..23354))
/gene="M4E13.60"
CDS complement(join(20820..21014,21397..21499,21644..21723,
21807..21889,22003..22051,22140..22205,22752..22801,
22966..23090,23242..23354))
Query Match 0.74; Score 21; DB 8; Length 80346;
Best Local Similarity 100.04; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 591 caaatgcctcttcctcccg 611
|||||
Db 75913 CAATTCGCTCTTCCTCCG 75933
RESULT 32
ATT12J5 84499 bp DNA PLN 24-FEB-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone T12J5 (ESSAIT
project).
DEFINITION ALO35522
ACCESSION ALO35522
VERSION ALO35522.1 GI:4455339
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 84499)
REFERENCE Bevan, M., Murphy, G., Ridley, P., Hudson, S., Hobeisel, J., Mewes, H.W.,
Mayer, K.F.X. and Schueller, C.
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 84499)
REFERENCE EU Arabidopsis sequencing project.
AUTHORS Direct Submission
TITLE Submitted (22-FEB-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosome 4 can be
viewed at: <http://webvr.mips.biochem.mpg.de/proj/thal/>.
FEATURES
source location/Qualifiers
1..84499
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1..49380
/note="position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL022023; for sequence analysis and annotation please
refer to this entry"
misc_feature
gene 48992..51335
/gene="T12J5.10"
gene complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
CDS complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
/note="similarity to Human (H326) mRNA, Homo sapiens,
gb:006631"

gb:006631"
/codon_start=1
/product="putative protein"
/protein_id="CAB36721.1"
/db_xref="GI:4455340"
/translation="MSEKARRGRSNGFLHHPVVDWEFRREVGISSRNFENRFSASEN
LVLRLEIKKLEKHGCVNTVFSAEEDVLISSDDRVVLMOMQGNVLSFSGHA
NNFOAKFMPEDRPTIVTCAGMGMPLRTEAPTELETCRSVDPRRNRMDAIDNLNIA
IDPNSNLFPAVGCEFEYARLXDIRRROGEGNLNFTRAARFCPPHLIGNDVGITGLA
FSESELVSYNDEFTLTTPGMGLSNLIPSSPISKSPVSKSESSSPADENHSVS
LVYKGNHCETVAGVNFEGPRSEYVSGDCGRIFLWRKKGGLIRYMEADENHVC
EPHPIHVLVASGLIESDIKVTSKAERARLPNIELPSRFRIPMISFSLFYDYDE
LEFGMDIGIDNGEGEESIDDAEDNDDSDVSSGVLDLNDSDDDMDSDDDVDVDM
DSDECFEIDNNMNMNGDGSFTVNDTSLSGHQDDVDVDD"
complement(48992..49360)
/gene="T12J5.10"
/number=1
exon complement(49361..49761)
/number=1
intron complement(49762..50493)
/gene="T12J5.10"
/number=2
exon complement(50494..50814)
/number=2
intron complement(50815..51066)
/gene="T12J5.10"
/number=3
exon complement(51067..51206)
/number=3
intron complement(51207..51335)
/gene="T12J5.10"
/number=4
exon complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
complement(51806..52129)
/gene="T12J5.20"
/number=1
exon 51806..53389
/gene="T12J5.20"
complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:039301"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36722.1"
/db_xref="GI:4455341"
/translation="MEESKRNLDDEKASLIDIMRYVFGFADIAAKCAIDLKIPPAI
ENHPSOPVLTSELSAVSASPSHLRIMRFLVHOGLEKFEVPTKDLGATGYTTPLSR
RMMTIKHGDLMAPADONLCHSOLINEAMACDARVVPVAGACOGLPDGVATVYDV
GGGTGEMGLIVKEFPRIKGFNFDLPHVIEAVYLDGVENVEGDMFDSIPASDAVIK
WVLHWGDKDCIKILNCKREAVLPNIGKVLIVCVIGEKNTMIAERDCKLEHRLQ
LDWVMVHSTGERTLKEWDFVLTEAGFAREVRPFDVDSILIAVRS"
complement(52130..52610)
/number=1
intron complement(52611..52939)
/gene="T12J5.20"
/number=2
exon complement(52940..53064)
/number=2
intron complement(53065..53389)
/gene="T12J5.20"
/number=3
exon 53847..57666
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/gene="T12J5.30"
CDS complement(join(55847..56170,56842..57666))
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:039301
contains EST gb:N96576, W43582"
/codon_start=1

Query Match	0.7%	Score 21:	DB 8:	Length 84499:
Best Local Similarity	100.0%	Pred. No. 22:		
Matches 21:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				
/number=4				
complement(64870, .64974)				
/number=4				
complement(64975, .65243)				
exon				
intron				
exon				
complement(64552, .64735)				
/number=3				
complement(64736, .64865)				
/gene="T1205, .50"				

NBPXRI.FNFNDICSRSGCKXKVPWVITPMDLLAQITMKNKHELSHVAVVSSSIDAPRL
 HPVGVLDRECIULTFR"
 intron
 /gene="F23EI2.10"
 /number=1
 456. .919
 /gene="F23EI2.10"
 /number=2
 920. .1088
 /gene="F23EI2.10"
 /number=2
 1089. 1610
 /gene="F23EI2.10"
 /number=3
 1611. .1677
 /gene="F23EI2.10"
 /number=3
 1678. 1890
 /gene="F23EI2.10"
 /number=4
 3124. .3241
 /gene="F23EI2.20"
 /number=1
 3124. .4428
 /gene="F23EI2.20"
 /number=1
 4211.3124. .3241.3323. .3465.3595. .3696.3945. .4107. .4428)
 /note="Similarity to dihydroflavonol-4-reductase, *Gerbera x sp.*, PIR2:S35189"
 /codon_start=1
 /product="putative protein"
 /protein_id="CA18727.1"
 /db_xref="GI:3080408"
 /db_xref="SPTREMBL:O65487"
 /translation="MDQKKKVCYVGTGASGFLASNLVYKRLLEGTEVYICVYRDQENKK
 LAHLKLEGAERLRIVKADLMEEGSPFNALMGQGVFHTASPEILRPALIEGLTNVLPR
 SCRRKPSLRVLYSSSTWYALSKTLAEQAAWKFSSENGIDLVTLPSFLVPSLP
 PDLCSTADVDGLLKGTEKFGMGOMGVHIDVDARTHIYFEEHAAGQYICSSNV
 ISLEELVEFLSRVSLPIPKRPDPLN"
 intron
 3242. .3322
 /gene="F23EI2.20"
 /number=1
 3323. .3465
 /gene="F23EI2.20"
 /number=2
 3466. .3594
 /gene="F23EI2.20"
 /number=2
 3595. .3696
 /gene="F23EI2.20"
 /number=3
 3697. .3944
 /gene="F23EI2.20"
 /number=3
 3945. .4107
 /gene="F23EI2.20"
 /number=4
 4108. .4210
 /gene="F23EI2.20"
 /number=4
 4211. .4428
 /gene="F23EI2.20"
 /number=5
 complement(join(4939. .5013.5109. .5195.5307. .5342, 5451. .5651.5725. .5811.6161. .6163))
 /gene="F23EI2.30"
 4939. .6163
 /gene="F23EI2.30"
 complement(join(4939. .5013.5109. .5195.5307. .5342, 5451. .5651.5725. .5811.6161. .6163))
 /gene="F23EI2.30"
 /note="Contains Clathrin adaptor complexes small chain

```

signature [VVYKRYASLYR]
contains EST gp:"T44219"
/codon_start=1
/product="clathrin assembly protein Ap19 homolog"
/protein_id="CA18728.1"
/db_xref="GI:3080409"
/db_xref="SPTREMBL:O23685"
/translation="MIHFVLLVSRQGVRLTKWYSPYAKERSKVTRELSCVILLNRGR
KICNVEVMGVYVVYKRYASLYFCMCIDQEDNELEVLIIHHVEILDRYGSCELLD
LIFNHKAVYIIDLLELIGELQESSKTVARIISQDLVEAKNEEASSISNIIAQMT
NR"
exon
    complement(4939..5013)
    /gene="F23E12.30"
    /number=1
    complement(5014..5108)
    /number=1
    complement(5109..5195)
    /gene="F23E12.30"
    /number=2
    complement(5196..5306)
    /number=2
    complement(5307..5342)
    /gene="F23E12.30"
    /number=3
    complement(5343..5450)
    /number=3
    complement(5451..5651)
    /gene="F23E12.30"
    /number=4
    complement(5652..5724)
    /number=4
    complement(5725..5811)
    /gene="F23E12.30"
    /number=5
    complement(5812..6160)
    /number=5
    complement(6161..6163)
    /gene="F23E12.30"
    /number=6
    complement(7352..7582)
    /gene="F23E12.40"
    /number=1
    7352..7582
    /gene="F23E12.40"
    complement(7352..7582)
    /gene="F23E12.40"
    complement(7352..7582)
    /gene="F23E12.40"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="CA18729.1"
    /db_xref="GI:3080410"
    /db_xref="SPTREMBL:O65488"
    /translation="MPLVRKNGVVRIGSPMLIHSENLVGDVEAPPGRGDEIGEEV
GGEIVTVSESCGEGCAQDAKTPPEGADRI"
    8233..9045
    /gene="F23E12.50"
    /number=1
    complement(8233..9045)
    /gene="F23E12.50"
    /note="similarity to DNA-binding protein PDI, Pisum
sativum"
    /codon_start=1
    /product="putative protein"
    /protein_id="CA18730.1"
    /db_xref="GI:3080411"
    /db_xref="SPTREMBL:O65489"
    /translation="MELNRSADAEKAKETPTGATSGATASGSSSGRRPRGRAGSK
NKPRLPTITRSPVLRSHLEVYSGSDISEAVSTATRRGCGVCIISGAGVANTWT

```

```

INOPAPAGGVITLHGRFDLISLNGTALPPAPAGGLTVYLAGGCGVYGVNAG
Query Match      0.7%; Score 21; DB 8; Length 86710;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 591 caaatcgctctctctccgcg 611
      |||||||
Db 11744 CAATTCGTCTCTCTCCGCG 11764

RESULT 34
AC004907      130117 bp      DNA      PRI      21-AUG-1999
LOCUS      Homo sapiens clone D10853H20, complete sequence.
DEFINITION      AC004907
ACCESSION      AC004907.2 GI:5757545
VERSION      HTG.
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 130117)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 130117)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (12-JUN-1998) Genome Sequencing Center, Washington
      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
      MO 63108, USA
      3 (bases 1 to 130117)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (21-AUG-1999) Genome Sequencing Center, Washington
      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
      MO 63108, USA
      On Aug 21, 1999 this sequence version replaced gi:3213090.
COMMENT      Location/Qualifiers
FEATURES
      source
      1..130117
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="D10853H20"

BASE COUNT      43826 a 22237 c 22107 g 41947 t
ORIGIN

Query Match      0.7%; Score 21; DB 9; Length 130117;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2109 cctcagttgctctcagcac 2129
      |||||||
Db 28683 CCTCAGTTGCTCTCAGCAC 28703

RESULT 35
AC079351      157308 bp      DNA      HTG      03-FEB-2001
LOCUS      Homo sapiens chromosome 7 clone RP11-828B13, WORKING DRAFT
DEFINITION      AC079351
ACCESSION      AC079351
VERSION      AC079351.3 GI:12658193
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 157308)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished

```

```

REFERENCE      2 (bases 1 to 157308)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (28-AUG-2000) Genome Sequencing Center, Washington
      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
      MO 63108, USA
      On Feb 3, 2001 this sequence version replaced gi:11415227.
COMMENT
      ----- Genome Center -----
      Center: Washington University Genome Sequencing Center
      Center code: WUGSC
      Web site: http://genome.wustl.edu/gsc/index.shtml
      ----- Project Information -----
      Center project name: H_NH0828B13
      ----- Summary Statistics -----
      Sequencing vector: p13; 93%
      Sequencing method: M13; 93%
      Chemistry: Dye-terminator ET; 93% of reads
      Chemistry: Dye-terminator Big Dye; 6% of reads
      Assembly program: Phrap; version 0.990319
      Consensus quality: 151471 bases at least Q40
      Consensus quality: 153455 bases at least Q30
      Consensus quality: 154315 bases at least Q20
      Insert size: 162000; agarose-fp
      Insert size: 156208; sum-of-ctnigs
      Quality coverage: 5.45 in Q20 bases; agarose-fp
      Quality coverage: 5.28 in Q20 bases; sum-of-ctnigs
      -----
      * NOTE: This is a 'working draft' sequence. It currently
      * consists of 12 contigs. The true order of the pieces
      * is not known and their order in this sequence record is
      * arbitrary. Gaps between the contigs are represented as
      * runs of N, but the exact sizes of the gaps are unknown.
      * This record will be updated with the finished sequence
      * as soon as it is available and the accession number will
      * be preserved.
      1
      6907      6906: contig of 6906 bp in length
      *
      7007      7006: gap of unknown length
      *
      16305      16304: contig of 9298 bp in length
      *
      16405      16404: gap of unknown length
      *
      16405      27392: contig of 10988 bp in length
      *
      27393      27492: gap of unknown length
      *
      27493      38821: contig of 11329 bp in length
      *
      38822      38921: gap of unknown length
      *
      38922      53583: contig of 14662 bp in length
      *
      53584      53683: gap of unknown length
      *
      53684      71646: contig of 17963 bp in length
      *
      71647      71746: gap of unknown length
      *
      71747      96546: contig of 24600 bp in length
      *
      96547      96646: gap of unknown length
      *
      96647      119634: contig of 22988 bp in length
      *
      119635      119734: gap of unknown length
      *
      119735      146540: contig of 26606 bp in length
      *
      146541      146640: gap of unknown length
      *
      146641      150320: contig of 3660 bp in length
      *
      150321      150420: gap of unknown length
      *
      150421      153662: contig of 3242 bp in length
      *
      153663      153762: gap of unknown length
      *
      153763      157308: contig of 3546 bp in length.
      FEATURES
      source
      1..157308
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="7"
      /clone="RP11-828B13"
      1..6906
      /note="assembly_name:Contig10"
      7007..16304
      /note="assembly_name:Contig11"
      16405..27392
      /note="assembly_name:Contig12"
      27493..38821
      /note="assembly_name:Contig13"
      misc_feature
      misc_feature
      misc_feature
      misc_feature

```



```
misc_feature      38922..53583
                  /note="assembly_name:Contig14
                  clone_end:17
                  vector_side:left"
misc_feature      53684..71646
                  /note="assembly_name:Contig15"
                  /note="assembly_name:Contig16"
                  /note="assembly_name:Contig17"
misc_feature      96647..119634
                  /note="assembly_name:Contig17"
misc_feature      119735..146540
                  /note="assembly_name:Contig18"
misc_feature      146641..150320
                  /note="assembly_name:Contig7"
misc_feature      150421..153662
                  /note="assembly_name:Contig8"
misc_feature      153763..157308
                  /note="assembly_name:Contig9
                  /note="assembly_name:Contig9
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      52313 a 26687 c 27184 g 50018 t 1106 others
ORIGIN
Query Match      0.7%; Score 21; DB 2; Length 157308;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2109 cctcagttgtcctcagcac 2129
      |||||||
      24594 CCTCAGTTGTGCTCCACAC 24614

RESULT 36
AC024590/c      183839 bp      DNA      HTG      25-APR-2001
LOCUS      Homo sapiens chromosome 16 clone RP11-487C14, WORKING DRAFT
AC024590
AC024590      14 unordered pieces.
AC024590      4 GI:13786378
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVERFIN.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 183839)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 183839)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint
      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
      On Apr 25, 2001 this sequence version replaced gi:9954669.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 593435
Center clone name: RPCI-11_487C14
-----
Summary Statistics
Consensus quality: 172462 bases at least Q40
Consensus quality: 177746 bases at least Q30
Consensus quality: 178816 bases at least Q20
Estimated insert size: 163610; agarose-ef estimation
Estimated insert size: 162539; sum-of-contigs estimation
Quality coverage: 14.49 in Q20 bases; agarose-ef estimation
Quality coverage: 12.99 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
```

```
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
      1 1928: contig of 1928 bp in length
      * 1929 2028: gap of unknown length
      * 2029 3128: contig of 1100 bp in length
      * 3129 3228: gap of unknown length
      * 3229 4300: contig of 1072 bp in length
      * 4301 4400: gap of unknown length
      * 4401 5574: contig of 1174 bp in length
      * 5575 5674: gap of unknown length
      * 5675 6732: contig of 1058 bp in length
      * 6733 6832: gap of unknown length
      * 6833 7977: contig of 1145 bp in length
      * 7978 8077: gap of unknown length
      * 8078 10923: contig of 2846 bp in length
      * 10924 11023: gap of unknown length
      * 11024 14086: contig of 3063 bp in length
      * 14087 14186: gap of unknown length
      * 14187 22792: contig of 8606 bp in length
      * 22793 22892: gap of unknown length
      * 22893 29405: contig of 6513 bp in length
      * 29406 29505: gap of unknown length
      * 29506 38983: contig of 9478 bp in length
      * 38984 39083: gap of unknown length
      * 39084 69212: contig of 30129 bp in length
      * 69213 69312: gap of unknown length
      * 69313 109646: contig of 40334 bp in length
      * 109647 109746: gap of unknown length
      * 109747 183839: contig of 74093 bp in length.
      Location/Qualifiers
        source
          1..183839
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="RP11-487C14"
            /clone_id="RPCI human BAC library 11"
BASE COUNT      50219 a 37425 c 39803 g 55084 t 1308 others
ORIGIN
Query Match      0.7%; Score 21; DB 2; Length 183839;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1977 cagcactctgaggtcacact 1997
      |||||||
      42601 CAGCACTCTGAGGTCACACT 42581

RESULT 37
AC013553/c      184864 bp      DNA      HTG      13-DEC-2000
LOCUS      Homo sapiens chromosome 15 clone RP11-325L12 map 15, *** SEQUENCING
AC013553
AC013553      14 ordered pieces.
AC013553      14 GI:11693406
HTG: HTGS_PHASE2.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 184864)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 15, clone RP11-325L12
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 184864)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
      Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
```

TITLE	Brown, A., Castelle, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Domingo, M., Donegan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Lehoczky, J., Lien, C., Locke, K., Macdonald, P., Marquis, N., McKwan, P., McGurk, A., McKennan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Sudramanlan, A., Talamas, J., Teefaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.			
COMMENT	Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Dec 13, 2000 this sequence version replaced gi:9280765. All repeats were identified using RepeatMasker: Smit, A.F.P. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html			
JOURNAL	----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence.submissions@genome.wi.mit.edu ----- Project Information Center project name: L2203 Center clone name: 325_L-12			
FEATURES	----- * NOTE: This is a 'working draft' sequence. It currently * consists of 1 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * provided by the submittor. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved. 1 184864: contig of 184864 bp in length. Location/Qualifiers 1..184864 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="15" /map="15" /clone="RP11-325L12" /clone_11b="RPC1-11 Human Male BAC" BASE COUNT 51274 a 42968 c 41143 g 49479 t			
ORIGIN				
Query Match	0.7%; Score 21; DB 2; Length 184864;			
Best Local Similarity	100.0%; Pred. No. 22;			
Matches 21; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
yy	259 caaagctcttttcacacagaca 279 			
Db	46943 CAAAGCTTTTTCATCAGACA 46923			
RESULT 38				
AC080090				
LOCUS	AC080090 194874 bp DNA HTG 26-Nov-2000			
DEFINITION	Homo sapiens chromosome 7 clone RP11-79616, WORKING DRAFT SEQUENCE, 29 unordered pieces.			
ACCESSION	AC080090			
VERSION	AC080090.3 GI:11345029			
KEYWORDS	HTG; HTGS_PHASEL; HTGS_DRAFT.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE -	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 194874) Waterston, R. H.			

```

TITLE      The sequence of Homo sapiens clone
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 194874)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL     Submitted (23-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis
            MO 63108, USA
COMMENT     On Nov 26, 2000 this sequence version replaced gi:11276326.

-----
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Project name:H_NH0796106
Project Information -----
Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178730 bases at least Q40
Consensus quality: 184228 bases at least Q30
Consensus quality: 186326 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 191473; sum-of-contigs
Quality coverage: 3.95 in Q20 bases; agarose-fp
Quality coverage: 4.25 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
*       1        601: contig of 601 bp in length
*         602          701: gap of unknown length
*         702          2533: contig of 1832 bp in length.
*         2534          2633: gap of unknown length
*         2634          3924: contig of 1291 bp in length
*         3925          4024: gap of unknown length
*         4025          5276: contig of 1252 bp in length
*         5277          5376: gap of unknown length
*         5377          6812: contig of 1436 bp in length
*         6813          6912: gap of unknown length
*         6913          9058: contig of 2146 bp in length
*         9059          9158: gap of unknown length
*         9159          11288: contig of 2130 bp in length
*         11289          13955: gap of unknown length
*         11389          13955: contig of 2567 bp in length
*         13956          14055: gap of unknown length
*         14056          16269: contig of 2214 bp in length
*         16270          16369: gap of unknown length
*         16370          19134: contig of 2765 bp in length
*         19135          19234: gap of unknown length
*         19235          22939: contig of 3705 bp in length
*         22940          23039: gap of unknown length
*         23040          26095: contig of 3056 bp in length
*         26096          26195: gap of unknown length
*         26196          29786: contig of 3591 bp in length
*         29787          29886: gap of unknown length
*         29887          32815          32914: contig of 2928 bp in length
*         32815          32915          37639: gap of unknown length
*         32915          37639: contig of 4725 bp in length
*         37640          37739: gap of unknown length
*         37740          41000: contig of 3361 bp in length
*         41001          41100: gap of unknown length
*         41101          46784: contig of 5684 bp in length
*         46785          46884: gap of unknown length
*         46885          54478: contig of 7594 bp in length

```

```
* 54479 54578: gap of unknown length
* 54579 61716: contig of 7138 bp in length
* 61716 61817: gap of unknown length
* 61817 68787: contig of 6970 bp in length
* 68787 68887: gap of unknown length
* 68887 76504: contig of 7618 bp in length
* 76504 76605: gap of unknown length
* 76605 86157: contig of 9553 bp in length
* 86157 86258: gap of unknown length
* 86258 95154: contig of 8897 bp in length
* 95154 95255: gap of unknown length
* 95255 105637: contig of 10383 bp in length
* 105637 105738: gap of unknown length
* 105738 119881: contig of 14143 bp in length
* 119881 131202: gap of unknown length
* 131202 131303: contig of 11222 bp in length
* 131303 144326: gap of unknown length
* 144326 144427: contig of 13024 bp in length
* 144427 164332: gap of unknown length
* 164332 164433: contig of 19906 bp in length
* 164433 194874: gap of unknown length
* 194874 304442: contig of 304442 bp in length.
```

```
Location/Qualifiers
1. 194874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-79616"
```

```
misc_feature
/note="assembly_name:Contig10"
702..2533
/note="assembly_name:Contig19"
2634..3924
/note="assembly_name:Contig11"
4025..3276
/note="assembly_name:Contig14"
5377..6812
/note="assembly_name:Contig16"
6913..9058
/note="assembly_name:Contig17"
9159..11288
/note="assembly_name:Contig18"
11389..13955
/note="assembly_name:Contig20"
14056..16269
/note="assembly_name:Contig21"
16370..19134
/note="assembly_name:Contig22"
19235..22939
/note="assembly_name:Contig23"
23040..26095
/note="assembly_name:Contig24"
26196..29786
/note="assembly_name:Contig25"
29887..32814
/note="assembly_name:Contig26"
32915..37639
/note="assembly_name:Contig27"
37740..41000
/note="assembly_name:Contig28"
41101..46784
/note="assembly_name:Contig29"
46885..54478
/note="assembly_name:Contig30"
54579..61716
/note="assembly_name:Contig31"
61817..68786
/note="assembly_name:Contig32"
68887..76504
/note="assembly_name:Contig33"
76605..86157
/note="assembly_name:Contig34"
86258..95154
misc_feature
```

```
misc_feature
/note="assembly_name:Contig35"
95255..105637
/note="assembly_name:Contig36"
105738..119880
/note="assembly_name:Contig37"
119881..131202
/note="assembly_name:Contig38"
131303..144326
/note="assembly_name:Contig39"
144427..164332
/note="assembly_name:Contig40"
164433..194874
/note="assembly_name:Contig41"
194874..304442
misc_feature
BASE COUNT 63657 a 32523 c 33628 g 62260 t 2806 others
ORIGIN
```

```
Query Match 0.7%; Score 21; DB 2; Length 194874;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2109 cctcagttgtcctcagcac 2129
|||||
Db 52703 CCTCAGTTGTGCTCCAGCAC 52723
```

```
RESULT 39
ATCHRIV82 195165 bp DNA PLN 16-MAR-2000
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 82.
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 82.
ACCESSION AL161586
VERSION AL161586.2 GI:7270418
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 29459 to 132608; 129923 to 195165)
Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
2 (bases 98968 to 179313)
Purnelle,B., Boutry,M., Goffeau,A., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
3 (bases 1 to 44270)
Terry,N., Ardiles,W., Buyshaert,C., Dasseville,R., De Clerck,R.,
De Keyser,A., Neyt,P., Rouze,P., Van Den Daele,H., Villarroel,R.,
Glezen,J., Van Montagu,M., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
4 (bases 1 to 195165)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
COORDINATOR: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV81 at the 5' end and an
overlap with ATCHRIV83 at the 3' end.
Location/Qualifiers
1. 195165
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
8478..8759
/gene="AT4g34680"
```

```

gene      /number=1
8478. .9388
/gene="AT4g34680"
join(8478. .8759,8861. .9388)
/gene="AT4g34680"
/note="contains EST gb:A1994545.1, R30028"
/codon_start=1
/product="GATA transcription factor 3"
/protein_id="CAB80185.1"
/db_xref="GI:7270419"
/translation="MEIWEARALKASLUGESTISLKHQVIVSEDIARTSSLPDEF
VCEPLDSGCKEEVEVSVSSQOEQHDVCSQCLFDOLPSPDDVPELE
WYSRVVDCSSPEVSLILQTHKTKPSFRITPKPKRNSLTGSRWPLVSTNHQ
HATQDLRRKKQETVLFQRCSHCGTNTPQWRTPGVPKTLCAACGVRFKSGRLCP
HATPADSPFSPNEIHSNLHRKYLELRKSKELGEETGEASTKSDPVKFGSKW"
8760. .8860
/gene="AT4g34680"

Intron   /number=1
8861. .9388
/gene="AT4g34680"

exon     /number=2
10241. .10875
/gene="AT4g34690"
complement(10241. .10855)
/gene="AT4g34690"

gene      /number=1
complement(join(10241. .10855,10867. .10875))
/gene="AT4g34690"
complement(join(10241. .10855,10867. .10875))
/gene="AT4g34690"

CDS      /codon_start=1
/product="hypothetical protein"
/protein_id="CAB80186.1"
/db_xref="GI:7270420"
/translation="MDLSDDGSGEKKSNDSGDKKSDGSSNNDDDDYTPNDPOS
TPEKVKLATRLHYIGVRSCEGRSYKRWKRLKFINISQAKKLOLLAKRKHDPF
VETCLGKASRDPSTYTSFTLWCFEGSLTCSFWGLWNSTSDRSLDLWLEINRK
SVVHIEFKEMSAKNALYMRIDETTKVAKQNOEDLOKIVALLVSK"
complement(10856. .10866)
/number=1
complement(10867. .10875)
/gene="AT4g34690"

exon     /number=2
11652. .13140
/gene="AT4g34700"
join(11652. .11783,12719. .12800,12946. .13001,13099. .13140)
/gene="AT4g34700"
/note="contains EST gb:T04342, T13747, AA040992,
A1992559.1, Z33673, H36271, T43433, T43941"
/codon_start=1
/product="putative protein"
/protein_id="CAB80187.1"
/db_xref="GI:7270421"
/translation="MSGVSTAAVFAFARAOKERVRLVRRALKDTLNMVAHHIIFRD
EVDRLDKLIANGEALEYKWRHPDPIYVMAFGSGFCFKNPPACIETIVYNGLEDN
P"
11652. .11783
/gene="AT4g34700"

exon     /number=1
11784. .12718
/gene="AT4g34700"

Intron   /number=1
12719. .12800
/gene="AT4g34700"

exon     /number=2
12801. .12945
/gene="AT4g34700"

Intron   /number=2
12946. .13001
/gene="AT4g34700"

exon     /number=3
13002. .13098
/gene="AT4g34700"

Intron

```

```

exon     /number=3
13099. .13140
/gene="AT4g34700"

gene      /number=4
complement(15093. .17228)
/gene="AT4g34710"
complement(15093. .17228)
/gene="AT4g34710"
/note="Contains Orn/DAP/Arg decarboxylases family 2
signatures AA144-162;Orn/DAP/Arg decarboxylases family 2
signatures AA325-338
contains EST gb:W43783, Z33969, Z33688, T46784,
A1998693.1, AA04845"
/codon_start=1
/product="arginine decarboxylase SPE2"
/protein_id="CAB80188.1"
/db_xref="GI:7270422"
/translation="MPALACVDFSPVPAYAFSDPAQGVFIASSPTSAAYVDRMSP
SLSSSLRIDGAPFTANSNGNTSVPHGSETLPHODDIDLKTVKRYGPKSSGGL
GLQPLIVAFPDVILKNRLECLQSAFDVAIKSQGYDSHYQGVYPAKCNDRVVEDIVK
FGSFRFGLEAGSKPEILILAMSLCKGSPDAFLVYNGFRDAEYISLALGKRLANTV
ILVEQEEELDIYIELSOKMNVPRVIGRAKLRTKSHGFGSGTSGKFGGLTTQIVR
VYRKLRQSGMLDCLQLLHPHIGSOIIPSTSLSDGVAEAALCYCELVRGAAHKYVDIG
GGLGTDYGSKSGESDLSVAYSLEERAAVAVSVVCDRSVKRPVYCSEGRATVS
HHSVLIFEAVSADKPMVHOATPGDIQLEEGNEARANTEDLTAVMKGDHESCLIV
DOLKORCVGEKEGYLSIEOLASVDGLCEWYKAIGASDPVHTYINISVFTSIPDLW
GIDOLFPIVPIHKLDQRPARGILSDLCDSGKINKFTIGSSSLPHELDKNGSGGR
YFLGMEIGAYEEALGVNLEGPSSVYRVSDGPHSFAYTRAVPAGQSSADYLRAMQ
HEPELMFOTLKHRAEEMHTKGSNGSENEEEDFENNVAASLDRSHNNPYLATQD
ASPSNLSAISNLGFTYCDEVDYDTISA"
complement(15093. .17228)
/gene="AT4g34710"

exon     /number=1
15093. .17228
/gene="AT4g34710"
complement(join(23001. .23133,23231. .23516,23868. .23943))
/gene="AT4g34720"
complement(23001. .23133)
/gene="AT4g34720"

gene      /number=1
23001. .23943
/gene="AT4g34720"
complement(join(23001. .23133,23231. .23516,23868. .23943))
/gene="AT4g34720"
/note="Contains Prokaryotic membrane lipoprotein lipid
attachment site AA16-26;Prokaryotic membrane lipoprotein
lipid attachment site AA91-101
contains EST gb:T41774, N96695, N38619, Z26112, N38618,
A1998496.1, A1996086.1, R90087, AA12128, H76763"
/codon_start=1
/product="vacuolar H+-transporting ATPase 16k chain"
/protein_id="CAB80189.1"
/db_xref="GI:7270423"
/translation="MSTSGDETAFFPGFLGAALVSCMAAYCAKSGVNASMG
VKRPELVKMSIVPYVMAGVLTIGLIIAVITSTGINPAKSTYLEDGTAHSSGLACG
IAGLSAGMAIGIVGAGVANAQPKFLVGMILLIFAEALALVGLIVGILISRAQ
SRAE"
complement(23134. .23230)
/number=1
complement(23231. .23516)
/gene="AT4g34720"

Intron   /number=2
23517. .23867
complement(23868. .23943)
/gene="AT4g34720"

exon     /number=3
24797. .25947
/gene="AT4g34730"
complement(join(24797. .24955,25226. .25399,25678. .25947))
/gene="AT4g34730"
complement(join(24797. .24955,25226. .25399,25678. .25947))
/gene="AT4g34730"

CDS

```

/note="similarity to similarity predicted protein,
 Synecocystis sp., PR2:S76285
 contains EST gb:f19817"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80190.1"
 /db_xref="GI:7270424"
 /translation="MAPVNLPTNLVSRYTCVCMANPRVKNVAQVRELSDMLLTPD
 VLQHAVPEALGLADRLISLTITSDVSTNSNDLOVMRPNKLLCNVKKYVSVFQDD

Query Match 0.7%; Score 21; DB 8; Length 195165;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 591 caaatcgctctctctccgcg 611
 |||||||
 Db 174880 CAATTCCTCTTCTCTCCGCG 174900

RESULT 40
 AC025566/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC025566 195217 bp DNA HTG 24-AUG-2000
 Homo sapiens chromosome 3 clone RP11-50102, WORKING DRAFT SEQUENCE,
 40 unordered places.
 AC025566
 AC025566.13 GI:9885893
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 195217)
 Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
 Bodoia,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
 Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
 David,R., Delgado,O., Deshazo,D., Ding,Y., Donah-Rashid,N.,
 Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
 Forcuna-Tansey,J., Frantz,P., Ganssh,R., Gorrell,J.H., Gorrell,L.L.,
 Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hughes,M.,
 Holloway,C., Hosack,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
 Kelly,S., Kondolewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
 Lichtege,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
 Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
 Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,S.,
 Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
 Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
 Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Suqang,R.,
 Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabhdh,M.,
 Wallington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
 Wolley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
 Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 195217)
 Worley,K.C.
 Direct Submission
 Submitted (11-MAR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Aug 23, 2000 this sequence version replaced gi:9802618.

COMMENT
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HAUJ
 Center clone name: RP11-50102
 ----- Summary Statistics
 Sequencing vector: M13; 108821
 Chemistry: Dye-primer; Bodyfig: 17% of reads
 Chemistry: Dye-terminator; Big Dye: 83% of reads
 Assembly program: Phrap; version 0.990329

Consensus quality: 151852 bases at least Q40	
Consensus quality: 165851 bases at least Q30	
Consensus quality: 114484 bases at least Q20	
Estimated insert size: 175615; sum-of-coverage estimation	
Quality coverage: 0x in Q20 bases; agorose-1p estimation	
Quality coverage: 3.1x in Q20 bases; sum-of-coverage estimation	

NOTE: Estimated insert size may differ from sequence length	
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).	
NOTE: This is a 'working draft' sequence. It currently	
consists of 40 contigs. The true order of the pieces	
is not known and their order in this sequence record is	
arbitrary. Gaps between the contigs are represented as	
runs of N, but the exact sizes of the gaps are unknown.	
This record will be updated with the finished sequence	
as soon as it is available and the accession number will	
be preserved.	
1	17628: contig of 17628 bp in length
17629	17728: gap of unknown length
17729	30651: contig of 12933 bp in length
30652	30751: gap of unknown length
30752	39810: contig of 9059 bp in length
39811	39910: gap of unknown length
39911	47904: contig of 7994 bp in length
47905	48004: gap of unknown length
48005	56310: contig of 8306 bp in length
56311	56410: gap of unknown length
56411	63568: contig of 7158 bp in length
63569	66668: gap of unknown length
66669	70324: contig of 6656 bp in length
70325	70424: gap of unknown length
70425	77997: contig of 7573 bp in length
77998	78097: gap of unknown length
78098	85133: contig of 7036 bp in length
85134	85233: gap of unknown length
85234	91183: contig of 5950 bp in length
91184	91283: gap of unknown length
91284	95667: contig of 4384 bp in length
95668	95767: gap of unknown length
95768	101736: contig of 5969 bp in length
101737	101836: gap of unknown length
101837	108638: contig of 6802 bp in length
108639	108738: gap of unknown length
108739	114324: contig of 5586 bp in length
114325	114424: gap of unknown length
114425	118727: contig of 4303 bp in length
118728	118827: gap of unknown length
118828	122769: contig of 3942 bp in length
122770	122869: gap of unknown length
122870	129332: contig of 6463 bp in length
129333	129432: gap of unknown length
129433	134166: contig of 4734 bp in length
134167	134266: gap of unknown length
134267	139324: contig of 5058 bp in length
139325	139424: gap of unknown length
139425	144289: contig of 4865 bp in length
144290	144389: gap of unknown length
144390	149162: contig of 4773 bp in length
149163	149262: gap of unknown length
149263	152931: contig of 3665 bp in length
152932	153031: gap of unknown length
153032	157664: contig of 4633 bp in length
157665	157764: gap of unknown length
157765	163426: contig of 5662 bp in length
163427	163526: gap of unknown length
163527	166803: contig of 3277 bp in length
166804	166903: gap of unknown length
166904	169178: contig of 2275 bp in length
169179	169278: gap of unknown length
169279	172648: contig of 3370 bp in length
172649	172748: gap of unknown length
172749	175179: contig of 2431 bp in length
175180	175279: gap of unknown length

```

* 175280 177372: contig of 2093 bp in length
* 177373 177472: gap of unknown length
* 177473 179336: contig of 1864 bp in length
* 179337 179436: gap of unknown length
* 179437 180528: contig of 1092 bp in length
* 180529 180628: gap of unknown length
* 180629 182229: contig of 1601 bp in length
* 182230 182329: gap of unknown length
* 182330 184954: contig of 2625 bp in length
* 184955 185054: gap of unknown length
* 185055 186495: contig of 1441 bp in length
* 186496 186595: gap of unknown length
* 186596 188317: contig of 1722 bp in length
* 188318 188417: gap of unknown length
* 188418 189682: contig of 1265 bp in length
* 189683 189782: gap of unknown length
* 189783 191715: contig of 1933 bp in length
* 191716 191815: gap of unknown length
* 191816 192922: contig of 1107 bp in length
* 192923 193022: gap of unknown length
* 193023 194102: contig of 1080 bp in length
* 194103 194202: gap of unknown length
* 194203 195217: contig of 1015 bp in length.

FEATURES
    source
        1..195217
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-50102"

BASE COUNT      61010 a 36076 c 34593 g 59532 t 4006 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1111 atacagaanaaacacacaaa 1131
Db 188931 ATACAGAAAAACAAACAAA 188911

RESULT 41
ATCHRIV83/c ATCHRIV83 197859 bp DNA PLN 16-MAR-2000
LOCUS
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.
ACCESSION AL161587
VERSION AL161587.2 GI:7270470
KEYWORDS
SOURCE
    ORGANISM
        Arabidopsis thaliana
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
        Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
    REFERENCE
        1 (bases 16641 to 103350)
            Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
            Unpublished
    JOURNAL
        2 (bases 98352 to 125759)
            Robben,J., Grymoprez,B., Volckaert,G. Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
            Unpublished
    JOURNAL
        3 (bases 120761 to 197859)
            Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
            Unpublished
    JOURNAL
        4 (bases 1 to 24256)
            Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
            Unpublished
    JOURNAL
        5 (bases 1 to 197859)
            EU Arabidopsis sequencing,project.
            Direct Submission

```

```

JOURNAL
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/tha1/
this fragment has an overlap with ATCHRIV82 at the 5' end and an
overlap with ATCHRIV84 at the 3' end.

COMMENT
FEATURES
    source
        1..197859
            /organism="Arabidopsis thaliana"
            /variety="Columbia"
            /db_xref="taxon:3702"
            /chromosome="4"
            11339..12705
                /gene="AT4g35190"
                join(11339..11494,11536..11664,11853..11952,12119..12189,
                12314..12457,12499..12705)
                /gene="AT4g35190"
                /note="similarity to ctf, Mycobacterium leprae, gb:U15180
                contains EST gb:T45691, T22640, T75954"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB80236.1"
                /db_xref="GI:7270471"
                /translation="MEIYKSRFRKRCVRCGSSGKRECYSDAATDIADELVRLCLNLM
                ESENLKWTIRLNLVYGSGTGLMGLVSAVHEHGVGLVIAQITDIFLTITGETTG
                EVIAVADHMERKAEMARHSDFIALPLGSGYGLLEELVIAQAQGLHDKPGLLVNVDG
                YNNYLTFIDKAVDDGFTKPSQRHIFVSAPNAKELVOKLEIMKINKDKKDSIDLL
                FFDIQAKVPVNDGVIKSRMEVEKKVQPOOQOVVCSMTSMQTEIAL"
                11339..11494
                    /gene="AT4g35190"
                    /number=1
                11495..11535
                    /gene="AT4g35190"
                    /number=1
                11536..11664
                    /gene="AT4g35190"
                    /number=2
                11665..11852
                    /gene="AT4g35190"
                    /number=2
                11853..11952
                    /gene="AT4g35190"
                    /number=3
                11953..12118
                    /gene="AT4g35190"
                    /number=3
                12119..12189
                    /gene="AT4g35190"
                    /number=4
                12190..12313
                    /gene="AT4g35190"
                    /number=4
                12314..12457
                    /gene="AT4g35190"
                    /number=5
                12458..12498
                    /gene="AT4g35190"
                    /number=5
                12499..12705
                    /gene="AT4g35190"
                    /number=6
                13757..14518
                    /gene="AT4g35200"
                    complement(13757..14518)
                    /gene="AT4g35200"
                    complement(13757..14518)
                /note="similarity to various predicted proteins,

```

Arabidopsis thaliana
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80237.1"
 /db_xref="GI:7270472"
 /translation="MAVSEFVRSNSYPSPRQPAHAHDEQUTRLRSSDSSASSICOR
 LSNQDLHDSLEKMRILSLSTNLALSQDIKELDGLSLIDLCNIAKDAISQMEGLM
 EIOSILRRKPGDLGGEVKYLVSRFLKSLQKYLKSLQKQSKSTVASLVFGRAE
 AVTMAIPESLSEFMSGKACGKMSLVKMSQNKVTCAEANEFTRIDSEFQSKSLD
 MEDVNLSESCIQDIEDGIESLSKSLIKRVSILNI"
 complement(13757..14518)
 /gene="AT4g35200"
 /number=1
 16043..16795
 /gene="AT4g35210"
 16043..16795
 /gene="AT4g35210"
 /number=1
 16043..16795
 /gene="AT4g35210"
 /note="similarity to various predicted proteins,
 Arabidopsis thaliana"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80238.1"
 /db_xref="GI:7270473"
 /translation="MAVSEFVRSNSYPSPRQPAHAHDEQUTRLRSSDSSASSICOR
 LSNQDLHDSLEKMRILSLSTNLALSQDIKELDGLSLIDLCNIAKDAISQMEGLM
 EIOSILRRKPGDLGGEVKYLVSRFLKSLQKYLKSLQKQSKSTVASLVFGRAE
 AVTMAIPESLSEFMSGKACGKMSLVKMSQNKVTCAEANEFTRIDSEFQSKSLD
 VONLEICIQDLEDGIGLSKSLIKRVSILNI"
 16640..24256
 /note="position 76883-84499 overlaps to BAC clone F23E12,
 EMBL acc:AL022604; for sequence analysis please refer to
 this accession"
 17252..17657
 /gene="AT4g35220"
 /number=1
 17252..18625
 /gene="AT4g35220"
 join(17252..17657,17740..17784,18534..18625)
 /gene="AT4g35220"
 /note="similarity to predicted protein, Arabidopsis
 thaliana
 contains EST gb:F14002, T46305, T43454, A1996658.1"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80239.1"
 /db_xref="GI:7270474"
 /translation="MAVSEFVRSNSYPSPRQPAHAHDEQUTRLRSSDSSASSICOR
 LKPIRREYNGKGIYDISHRTPEKPSDSEGIQRLMLASKNGSLANSEMKIP
 THGTIVHDSPEGHVYDKYDAGFDVDSLQVNLALIVDPKDKNITDWLEQKLOP
 VASSSIDSLPQTCQIRNCP"
 17658..17739
 /gene="AT4g35220"
 /number=1
 17740..17784
 /gene="AT4g35220"
 /number=2
 17785..18533
 /gene="AT4g35220"
 /number=2
 18534..18625
 /gene="AT4g35220"
 /number=3
 19941..22657
 /gene="AT4g35230"
 complement(join(19941..20231,20345..20458,20546..20710,
 20782..20980,21340..21425,21551..21656,21755..21888,
 22078..22213,22304..22657))
 /gene="AT4g35230"
 complement(join(19941..20231,20345..20458,20546..20710,
 20792..20980,21340..21425,21551..21656,21755..21888,

22078..22213,22304..22657))
 /gene="AT4g35230"
 /note="similarity to protein kinase APK1, Arabidopsis
 thaliana, PIR2:S28615
 contains EST gb:A199799.1, F13911"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80240.1"
 /db_xref="GI:7270475"
 /translation="MGCCQSLFSGDNPLKDGVPQPLSONNHGATTAONGSGAS
 GVGGGGGGIPSESEFADLKAATNNFSDNIVSGEAPMLVYKGRONRMTA
 VKFTKMAPEPKQFAEAMGVGLRIHRIANLIGVCCDGERLLVAFEPNPLAKH
 LFHENQTIEMAMRLRVGYIAEALDYCSTFGRPLVDHIANRYLPEDDNDPLRSCG
 LMKSRDGSYSTNLATPPEXLENGVTPESYVTSCTYLILDSKHPISALDM
 IRKNITLLMDSHLEGFSTEEATVVELASQCLQYEPREPKTKDLVATLAPLQKS
 DVPSYVMIGIKQEEAPSTPDRPLVSGEASRDALAIHQILMTYVRDDEGTNELS
 FOEWTOQMMDLDAKRGDQSFREKDKRTIDCSQMSYGVIFHFCGQFIDVGTWSP
 TVFGRRSLCYLLCDQPDPAALDAMQACVYPDMPTAFYMQSVALARLNMWTDADMNLN
 EAAOLEERORGRGS"
 complement(119941..20231)
 /gene="AT4g35230"
 /number=1
 complement(20232..20344)
 /number=1
 complement(20345..20458)
 exon
 intron
 Query Match 0.7%; Score 21; DB 8; Length 197859;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 591 caaatcgctctctctccg 611
 ||||||||||||||||||||
 Db 91607 CAATTCGTCCTTCCTCCG 91587
 RESULT 42
 AC068573/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 198935)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 15, clone RP11-54056
 Unpublished
 2 (bases 1 to 198935)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguski,M., Bouguelter,B., Brown,A., Burkett,G.,
 Campione,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S.,
 Collumore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,U.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardina,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehotzky,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,
 Meldrim,J., Meneses,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Olivar,V., Peterson,K., Pierre,N.,
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Turrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 19, 2001 this sequence version replaced gi:10280764.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L10276

Center clone name: 540_E-6

----- Summary Statistics

Sequencing vector: M13; M7815; 41% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 193533 bases at least Q40
Consensus quality: 195810 bases at least Q30
Consensus quality: 196898 bases at least Q20
Insert size: 166000; agarose-fp
Insert size: 197835; sum-of-ctrligs
Quality coverage: 10.6 in Q20 bases; agarose-fp
Quality coverage: 8.9 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 4882: contig of 4882 bp in length
* 4883 4982: gap of 100 bp
* 4883 5624: contig of 642 bp in length
* 5625 5724: gap of 100 bp
* 5725 6734: contig of 1010 bp in length
* 6735 6834: gap of 100 bp
* 6835 8393: contig of 1559 bp in length
* 8394 8493: gap of 100 bp
* 8494 11316: contig of 2823 bp in length
* 11317 11416: gap of 100 bp
* 11417 16171: contig of 4755 bp in length
* 16172 16271: gap of 100 bp
* 16272 30341: contig of 14070 bp in length
* 30342 30441: gap of 100 bp
* 30442 49473: contig of 19032 bp in length
* 49474 49573: gap of 100 bp
* 49574 75583: contig of 26010 bp in length
* 75584 75683: gap of 100 bp
* 75684 158422: contig of 82739 bp in length
* 158423 158522: gap of 100 bp
* 158523 180063: contig of 21541 bp in length
* 180064 180164: gap of 100 bp
* 180164 198935: contig of 18772 bp in length.

FEATURES
source

1.198935

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="15"

/map="15"

/clone="RP11-540E6"

/clone_1lb="RPCT-11 Human Male BAC"

1.4882

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

4983.5624

/note="assembly_fragment"

5725.6734

misc_feature
6835..8393
/note="assembly_fragment"
misc_feature
8494..11316
/note="assembly_fragment"
misc_feature
11417..16171
/note="assembly_fragment"
misc_feature
16272..30341
/note="assembly_fragment"
misc_feature
30442..49473
/note="assembly_fragment"
misc_feature
49574..75583
/note="assembly_fragment"
misc_feature
75684..158422
/note="assembly_fragment"
misc_feature
158523..180063
/note="assembly_fragment"
misc_feature
180164..198935
/note="assembly_fragment"
clone_end:T7
vector_side:right"

BASE COUNT 53896 a 46178 c 45676 g 52083 t 1102 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 22;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 caaagcttttcacagaca 279

Db 13730 CAAAGCTTTTCATCAGACA 13710

RESULT 43

AL593857/c

LOCUS

DEFINITION

Mus musculus chromosome 2 clone RP23-439H2, *** SEQUENCING IN

PROGRESS ***; In unordered pieces.

AL593857

AL593857.1 GI:14787341

ACCESSION

AL593857

KEYWORDS

HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

house mouse.

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 206137)

REFERENCE

Direct Submission

Submitted (11-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: BM439H2

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 204045 bases at least Q40

Consensus quality: 204532 bases at least Q30

Consensus quality: 204877 bases at least Q20

Insert size: 205337; sum-of-ctrligs

Insert size: 207948; 2.7% error; agarose-fp

Quality coverage: 10.60x in Q20 bases; sum-of-ctrligs Quality

coverage: 10.47x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence

Genome Center

Alonsookos, S.L., Amairungu, H.C., Are, J.R., Banks, T., Barberia, J., Benton, J., Bimague, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, L., Davis, C., Davy-Carroll, L., Dedering, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elstner, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frintz, P., Gebist, A., Gao, J., Garcia, A., Garner, T.,

Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kova, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, J. C., Lewis, L., Li, J., Li, Z., Lichtenberg, O., Lieu, C., Liu, J., Liu, W., Louie, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, A., Wleczek, R., Wood, S., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wood, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Unpublished
2 (bases 1 to 221341)
Worley, K. C.

Submitted (09-ADG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center Project name: HDGU
Center Clone name: RP11-500K7

----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 212184 bases at least Q40
Consensus quality: 217356 bases at least Q30
Consensus quality: 220392 bases at least Q20
Estimated insert size: 220263; sum-of-coverage
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 4.1x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length -----
(see <http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html>).
NOTE: This is a 'working draft' sequence. It currently
consists of 20 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 22546: contig of 22546 bp in length
22547 22646: gap of unknown length
22647 48410: contig of 25764 bp in length
48411 48510: gap of unknown length
48511 71072: contig of 22562 bp in length
71073 71172: gap of unknown length
71173 91596: contig of 20426 bp in length
91597 91698: gap of unknown length
91699 108834: contig of 17136 bp in length
108835 108934: gap of unknown length
108935 126517: contig of 17583 bp in length
126518 126617: gap of unknown length

126618 139129: contig of 12512 bp in length
139130 139229: gap of unknown length
139230 148601: contig of 9372 bp in length
148602 148701: gap of unknown length
148702 157335: contig of 8634 bp in length
157336 157435: gap of unknown length
157436 166202: contig of 8767 bp in length
166203 173060: gap of unknown length
173061 173160: gap of unknown length
173161 180928: contig of 7768 bp in length
180929 181028: gap of unknown length
181029 188269: contig of 7241 bp in length
188270 188369: gap of unknown length
188370 195914: contig of 7545 bp in length
195915 196014: gap of unknown length
196015 200892: contig of 4878 bp in length
200893 200992: gap of unknown length
200993 205927: contig of 4935 bp in length
205928 206027: gap of unknown length
206028 210855: contig of 4828 bp in length
210856 210955: gap of unknown length
210956 214834: contig of 3879 bp in length
214835 214934: gap of unknown length
214935 217876: contig of 2942 bp in length
217877 221341: contig of 3365 bp in length.

FEATURES
source
1. 221341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q"
/clone="RP11-500K7"

BASE COUNT 70894 a 38966 c 37158 g 72403 t 1920 others

ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 221341;
Best local similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1111 ataccagaataaacacacaa 1131
|||||
Db 41072 ATACGAGAAAACACACAAA 41052

RESULT 46
AC020727
LOCUS AC020727 227194 bp DNA HTG 10-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-500K7, WORKING DRAFT SEQUENCE,
20 unordered pieces.
AC020727
AC020727.5 GI:10048064
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 227194)
Waterston, R. H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 227194)
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 10, 2000 this sequence version replaced gi:7232174.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

web site: http://genome.mutl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0500K07

----- Summary Statistics -----
Sequencing vector: M13, 87%
Sequencing vector: plasmid, 13%
Chemistry: Dye-primer ET; 87% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 215874 bases at least Q40
Consensus quality: 219492 bases at least Q30
Consensus quality: 221591 bases at least Q20
Insert size: 231000; agarose-fp
Insert size: 226623; sum-of-contigs
Quality coverage: 4.11 in Q20 bases; agarose-fp
Quality coverage: 4.19 in Q20 bases; sum-of-contigs
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1447: contig of 1447 bp in length
* 1448 1547: gap of unknown length
* 1548 2609: contig of 1062 bp in length
* 2610 2709: gap of unknown length
* 2710 4301: contig of 1592 bp in length
* 4302 4401: gap of unknown length
* 4402 7988: contig of 3587 bp in length
* 7989 8088: gap of unknown length
* 8089 10498: contig of 2410 bp in length
* 10499 10598: gap of unknown length
* 10599 13494: contig of 2896 bp in length
* 13495 13594: gap of unknown length
* 13595 17577: contig of 3983 bp in length
* 17578 17677: gap of unknown length
* 17678 20847: contig of 3170 bp in length
* 20848 20947: gap of unknown length
* 20948 24832: contig of 3885 bp in length
* 24833 24933 24932: gap of unknown length
* 24933 29810: contig of 4878 bp in length
* 29811 37323: contig of 7413 bp in length
* 37324 37424 37423: gap of unknown length
* 37424 47954: contig of 10531 bp in length
* 47955 48054: gap of unknown length
* 48055 60562: contig of 12508 bp in length
* 60563 60662: gap of unknown length
* 60663 76734: contig of 16072 bp in length
* 76735 76834: gap of unknown length
* 76835 91807: contig of 15073 bp in length
* 91808 92007: gap of unknown length
* 92008 11826: contig of 19819 bp in length
* 11827 11926: gap of unknown length
* 11927 132932: contig of 21006 bp in length
* 132933 133032: gap of unknown length
* 133033 134508: contig of 21476 bp in length
* 134509 154608: gap of unknown length
* 154609 185758: contig of 31150 bp in length
* 185759 185858: gap of unknown length
* 185859 227194: contig of 41336 bp in length.

FEATURES
source
1. 227194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-500K7"
1. 1447
/note="Assembly_name:Contig6"
1548..2609

```

	misc_feature	/note="assembly_name:Contig7"	2710..4301	
	misc_feature	/note="assembly_name:Contig8"	4402..7988	
	misc_feature	/note="assembly_name:Contig9"	8089..10498	
	misc_feature	/note="assembly_name:Contig10"	10399..113494	
	misc_feature	/note="assembly_name:Contig11"	13595..17577	
	misc_feature	/note="assembly_name:Contig12"	17678..20847	
	misc_feature	/note="assembly_name:Contig13"	20948..24032	
	misc_feature	/note="assembly_name:Contig14 clone_end:SP6 vector_side:right"	24933..229810	
	misc_feature	/note="assembly_name:Contig15"	29911..37323	
	misc_feature	/note="assembly_name:Contig16"	37424..47954	
	misc_feature	/note="assembly_name:Contig17"	48055..60562	
	misc_feature	/note="assembly_name:Contig18"	60663..76734	
	misc_feature	/note="assembly_name:Contig19"	76835..91907	
	misc_feature	/note="assembly_name:Contig20"	92008..111826	
	misc_feature	/note="assembly_name:Contig21"	111927..132932	
	misc_feature	/note="assembly_name:Contig22"	133033..154508	
	misc_feature	/note="assembly_name:Contig23"	154609..185758	
	misc_feature	/note="assembly_name:Contig24"	185859..227194	
	misc_feature	/note="assembly_name:Contig25"		
BASE COUNT	71864 a 40078 c 38264 g 75058 t	1930 others		
ORIGIN				
Query Match	0.7%;	Score 21;	DB 2;	Length 227194;
Best Local Similarity	100.0%;	Pred. No. 22;		
Matches 21; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
OY 1111 atacgagaaaacaacacaa 1131				
Db 182182 ATACAGAAAACAAACACCAA 182202				
RESULT 47				
AEO03538	AE003538	303367 bp	DNA	INV 06-OCT-2000
DEFINITION	Drosophila melanogaster genomic scaffold 14200013386050 section 30			
VERSION	AE003538 AE002602			
ACCESSION	AE003538.2 GI:10727947			
KEYWORDS	HTG.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 303367)			
REFERENCE	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.C., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,T.H., Blazer,R.G., Champé,M., Pfeiffer,B.D., Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abriil,J.F., Agbayani,A., An,H.J.,			

Andrews-Pfannkuch, C., Baldwin, D., Ballew, R.M., Basu, A.,
 Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
 Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
 Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C.,
 Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
 Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
 Pablo, S., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
 Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
 Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,
 Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S.,
 Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorell, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
 Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kenison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
 Kravitz, S., Kuip, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mody, B.,
 Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Paclet, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
 Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Massarman, D.A.,
 Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
 Zaveri, J.S., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 Morley, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 303367)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submision
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7294470.

location/Qualifiers
 1. 303367
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="3L"
 join(<27833..27889,28689..29999,30090..30152,
 30225..>30848)
 /gene="CG11281"
 /product="CT31489"
 /db_xref="FLYBASE:FBan0011281"
 /db_xref="FLYBASE:FBgn0036347"
 <27833..>30848
 /gene="CG11281"
 /map="69F7-70A1"
 /db_xref="FLYBASE:FBan0011281"
 /db_xref="FLYBASE:FBgn0036347"
 join(27833..27889,28689..29999,30090..30152,30225..30848)
 /gene="CG11281"
 /note="CG11281 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0011281"
 /db_xref="FLYBASE:FBgn0036347"
 /protein_id="AAF49838.1"
 /translation="MORVPPSATLTSCNCTNSGSGSTRICQCLRPVETLLGLVGLF
 LKRLALNFTLGRFLVNGATDLVFIIFVAVTCTIMLSIPRAVILLIFALVGS
 RYLAIVAFRIIISGPIDLVENAGAVARFVCTVLYTLNLSITRDLDLAPETNLK
 HMGDVEIRHFEYELQEVLYDKAVENSDEIDEXGDKNTKPIYERMGRETSRNV
 SEINGGKELPTPAVQERFORNMNRCKHQLSGRACLEVFRNGYRCKCTTFPSMI
 AKAIQWYRVDICELDLPGNPKICDPSAVYVONGETVVELLKAEKLEFNSOIV
 VNEIKDEQFAKSLAERTGQAFKEDPERCKRIINRWGILIKTLCEMLRMVYS
 INIYVYLVNVERDNTITIKFKHVDPQRREQIDAILPLRYEKSVIDVDHIFSR
 HHSTTVIRKRSKIYVQDGEHEVRNIGSGVMARLLRTTMNFNIHEKVSYLSLNK

ECLEPNHVLPRKKMYVOLLIIILYIIVLYSTFFFLMRRCISFFYKREKORILFLY
 NMLRKLRSLEFLIHDADNLATRIQOQVWFLMRSCVAGMTIRHFFARTC
 MLCRGLEDSTFVCGNGLPYCDCAEDLNISVFCQGVYLFTRGAGSSSESVYTYRK
 E"

129707..131939
 /gene="caps"
 /note="Nucleotide sequence of the Celera sequence differs
 from the published sequence for this transcript."
 /product="CT31495"
 /db_xref="FLYBASE:FBan0011282"
 /db_xref="FLYBASE:FBgn0023095"
 <129707..>131939
 /gene="caps"
 /note="CG11282"
 /map="70A2-70A3"
 /db_xref="FLYBASE:FBan0011282"
 /db_xref="FLYBASE:FBgn0023095"
 129840..131462
 /gene="caps"
 /note="caps gene product; Nucleotide sequence of the
 Celera sequence differs from the published sequence for
 this transcript"
 /codon_start=1
 /db_xref="FLYBASE:FBan0011282"
 /db_xref="FLYBASE:FBgn0023095"
 /protein_id="AAF49837.1"
 /db_xref="GI:7294494"
 /translation="MSRRRAITMSIAPHIGQAFSICICICILVATLIPVALNCP
 NCCRCDDPLTMNCGEGTIDVPIPLNPAIQMLYKNNKTKTIDSMQFYADLTLD
 SFNDMLTTERSFATYHAKLQELHLHNKIGQVSNKTFGLSTVNLNGNLTAELEY
 RFSPVVKLAENLQGNRISHIDPALDGLNLRVLYDDNTLTVPBELTFQALHSL
 ABLYLCTNSFNTIPGAFODLKGFLRLDAGLHNI SGDALKGLVSLRFDLSNRL
 PAIPNAFORLRLQNLINGNDPEFVLSGAFSGRLRELBELTGAORLRVSGAFS
 GNTNHEHLNLSNKNLNELSIALGSLPLSTVYVKRANOLSLDGLVPMADLQVLDL
 SNPFECDCRLMLHLIYSRNAGQVATCAYPALRDLPLAHLAPLIGCGARGAA
 SKQALIGILVACALITTLVLYTCRRIRIEMKKGHSALGRKREYQKTFSDREYM
 SRPPGGGVHPAGGYPIYAGNSRMIPYTEL"
 join(125807..155924,155990..156094)
 /gene="AcP70A"
 /note="Nucleotide sequence of the Celera sequence differs
 from the published sequence for this transcript."
 /product="CT39049"
 /db_xref="FLYBASE:FBan0017673"
 /db_xref="FLYBASE:FBgn0003034"
 <155807..>156094
 /gene="AcP70A"
 /note="CG17673"
 /map="70A3-70A4"
 /db_xref="FLYBASE:FBan0017673"
 /db_xref="FLYBASE:FBgn0003034"
 join(155810..155924,155990..156042)
 /note="AcP70A"
 /note="AcP70A gene product; Nucleotide sequence of the
 Celera sequence differs from the published sequence for
 this transcript"
 /codon_start=1
 /db_xref="FLYBASE:FBan0017673"
 /db_xref="FLYBASE:FBgn0003034"
 /protein_id="AAF49836.1"
 /db_xref="GI:7294493"
 /translation="MKTIALFLVLCVGLGVQAMWPMNRKPTKFPIDSPDRDKWR
 LNLGPMWGGRC"
 <156744..>156902
 /gene="CG14114"
 /product="CT33710"
 /db_xref="FLYBASE:FBan004114"
 /db_xref="FLYBASE:FBgn0040815"
 /evidence="not experimental"
 <156744..>156902
 /gene="CG14114"
 /map="70A3-70A4"
 /db_xref="FLYBASE:FBan004114"
 /db_xref="FLYBASE:FBgn0040815"

CDS
/evidence=not_experimental
156744..156902
/gene="CG14114"
/note="CG14114 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014114"
/db_xref="FLYBASE:FBgn0040815"
/evidence=not_experimental
/protein_id="AA049835.1"
/db_xref="GI:7294492"
/translation="MSQESGDRSPFICVREGLESQMPNKEITMTTVLSSSPQOMKTM
QMWNTLGH"
join(<183391..183408,183461..>183691)
/gene="CG14113"
/product="CT33709"
/db_xref="FLYBASE:FBan0014113"
/db_xref="FLYBASE:FBgn0040814"
/evidence=not_experimental
/gene="CG14113"
/map="70A4-70A4"
/db_xref="FLYBASE:FBan0014113"
/db_xref="FLYBASE:FBgn0040814"
/evidence=not_experimental
join(183391..183408,183461..183691)
/gene="CG14113"
/note="CG14113 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014113"
/db_xref="FLYBASE:FBgn0040814"
/evidence=not_experimental
/protein_id="AA049834.1"
/db_xref="GI:7294491"
/translation="MSSEFEYTFEETDQANSPIREHDLPMSELQVALRYRIIRENK
SLSDQVIGIREPNVAEVAPEPSPDMNDKENKAVARBP"
Query Match 0.7%; Score 21; DB 3; Length 303367;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1631 aaattctgaagccataaca 1631
|||||
Db 162175 AAATTCGAAAGCCATAACA 162195
RESULT 48
AY027789 768 bp mRNA PRI 20-JUL-2001
LOCUS Homo sapiens CLANC (CLAN1) mRNA, complete cds.
DEFINITION
ACCESSION AY027789
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE Clan, a novel human ced-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 768)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
Location/Qualifiers
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"

CDS
/map="2p22-p21"
/tissue_type="lung"
1..768
/gene="CLAN1"
277..747
/gene="CLAN1"
/codon_start=1
/product="CLANC"
/protein_id="AAK14778.1"
/db_xref="GI:14324117"
/translation="MNFIDKNSRALIRMGNTYIKQITDLEFWNVNIREVNILICE
KVEDDAARGIITHMLKSGSCNIFLKSLEWNPLODILNGOSLFPOTSGDLDIA
ODKIDLTFTSEFLNFPPLGEDIIDIFMLKSTFTFEPVLRRDQHHRVEQLTVL"
BASE COUNT 218 a 157 c 180 g 213 t
ORIGIN
Query Match 0.7%; Score 20; DB 9; Length 768;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2331 agaagatgctataaactag 2350
|||||
Db 755 AGAAGATGCTATAAAACTAG 736
RESULT 49
AC078439
LOCUS AC078439 865 bp DNA HTG 02-AUG-2000
DEFINITION Giardia intestinalis clone N5017 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC078439
VERSION AC078439.1 GI:9654056
KEYWORDS HTG; HTGS-PHASED.
SOURCE Giardia intestinalis.
ORGANISM Giardia intestinalis
Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
REFERENCE 1 (bases 1 to 865)
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Giardia: a model for ancient eukaryotic genome analysis
JOURNAL Unpublished
AUTHORS Nixon,J., Morrison,H.G., McArthur,A.G., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2000) Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Marine Biological Laboratory, 7
Mbl Street, Woods Hole, MA 02543-1015, USA
COMMENT * NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 865: contig of 865 bp in length.
Location/Qualifiers
1..865
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="N5017"
BASE COUNT 232 a 218 c 189 g 226 t
ORIGIN
Query Match 0.7%; Score 20; DB 2; Length 865;
Best Local Similarity 100.0%; Pred. No. 89;

Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 705 gctgctgaagctgcggcaga 724
 |||||||
 Db 800 GCTGCTGAAGCTCGGCAGA 819

RESULT 50
 AC047412 1092 bp DNA HTG 13-APR-2000
 LOCUS Giardia intestinalis clone KI0936 strain WB-C6, LOW-PASS SEQUENCE
 DEFINITION SAMPLING.
 AC047412
 VERSION AC047412.1 GI:7548543
 KEYWORDS HTG: HTGS_PHASE0.
 SOURCE Giardia intestinalis.
 ORGANISM Giardia intestinalis
 Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
 REFERENCE 1 (bases 1 to 1092)
 AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Kim,U.,
 Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
 TITLE Giardia: a model for ancient eukaryotic genome analysis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1092)
 AUTHORS Kim,U., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q.,
 Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
 TITLE Direct Submission
 JOURNAL Submitted (13-APR-2000) Josephine Bay Paul Center for Comparative
 Molecular Biology and Evolution, Marine Biological Laboratory, 7
 MBL Street, Woods Hole, MA 02543-1015, USA
 COMMENT * NOTE: This record contains 1 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1 1092: contig of 1092 bp in length.
 * Location/Qualifiers
 1..1092
 /organism="Giardia intestinalis"
 /strain="WB-C6"
 /db_xref="taxon:5741"
 /clone="KI0936"

BASE COUNT 274 a 270 c 277 g 268 t 3 others

ORIGIN

Query Match 0.7%; Score 20; DB 2; Length 1092;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 705 gctgctgaagctgcggcaga 724
 |||||||
 Db 123 GCTGCTGAAGCTCGGCAGA 142

RESULT 51
 REOS3NSB 1198 bp ds-RNA VRL 03-AUG-1993
 LOCUS Reovirus serotype 2 S3 segment nonstructural protein (sigma-NS)
 DEFINITION RNA complete cds.
 ACCESSION M18390.1 GI:333757
 VERSION M18390.1 GI:333757
 KEYWORDS Sigma-NS nonstructural protein.
 SOURCE Reovirus serotype 2 (strain D5/Jones) viral RNA, passed in mouse L
 fibroblast cells.
 ORGANISM Mammalian orthoreovirus 2
 Viruses; dsRNA viruses; Reoviridae; Orthoreovirus; Mammalian

REFERENCE 1 (bases 1 to 1198)
 AUTHORS Wiener,J.R. and Joklik,W.K.
 TITLE Comparison of the reovirus serotype 1,2, and 3 S3 genome segments
 JOURNAL Virology 161, 332-339 (1987)
 MEDLINE 88072071
 FEATURES
 source Location/Qualifiers
 1..1198
 /organism="Mammalian orthoreovirus 2"
 /db_xref="taxon:10885"
 28..1128
 /note="nonstructural protein (sigma-NS)"
 /protein_id="AAA47282.1"
 /codon_start=1
 /db_xref="GI:333758"
 /translation="MASSSLRAISKIKRBDNGQGYCPNVMRSSVNTKYRVNVDYD
 IKTGFPSGCIANLRLPYAKRRRLGKRNLEIRARVDYLOTDLSDLCMPDPAPMTN
 VQASTMRRLVCDHFVVDVGLRVPMDDRSVSSSLARLFTMGAGLHITTEPAKRV
 PIMHLAADIDCMTEFALPYMTIVDGDVPAVPAETLPAERLLDGFEGYGLDISGCEV
 DNNRSAGDOSMDSRCINELTYAETAEACILKICLINCQFLEMDLAHNGFEL
 DKVQMIPESEVFPMASAFATIDVOCFRCCLIMDKNLKIDRMFTMLWRAGSDDA
 ISTSLTITSLDGRHWVADMDNENRILVFPARY"

BASE COUNT 288 a 272 c 306 g 332 t

ORIGIN

Query Match 0.7%; Score 20; DB 14; Length 1198;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2144 attctctcatggtggaagcc 2163
 |||||||
 Db 358 ATTCTCTCATGCTGGAAGCC 339

RESULT 52
 SCYBR141C 1688 bp DNA PLN 11-AUG-1997
 LOCUS S.cerevisiae chromosome II reading frame ORF YBR141C.
 DEFINITION Z36010 Y13134
 ACCESSION Z36010 Y13134
 VERSION Z36010.1 GI:536439
 KEYWORDS
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Saccharomycetes; Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 1688)
 AUTHORS Becam,A.M., Herdelt,C.J., Nasr,F., Stonimski,P.P. and Zagulski,M.
 JOURNAL Unpublished
 TITLE 2 (bases 1 to 1688)
 JOURNAL MIPS.
 REFERENCE Direct Submission
 Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
 European yeast chromosome II sequencing project. MIPS at the
 Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
 3 (bases 1 to 1688)
 REFERENCE Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Bactlet,M.C.,
 Barthe,C., Baur,A., Becam,A.M., Billeau,N., Boles,E., Brandt,T.,
 Bardin,M., Bruckner,M., Bussereau,F., Christiansen,C.,
 Contreras,R., Crouzet,M., Cziepluch,C., Demolis,N., Delaveau,T.,
 Dolgon,F., Domdey,H., Duesterhus,S., Dubois,E., Dujon,B., El
 Bakoury,M., Entlan,K.D., Feuerhann,M., Fiers,W., Fobo,G.M.,
 Fritz,C., Gassenhuber,H., Glansdorff,N., Goffeau,A., Griwall,L.A.,
 de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
 Jacq,C., Jaquet,M., Jauniaux,J.C., Joniaux,J.L., Kallioe,T.,
 Kiesau,P., Kirchath,L., Koelter,P., Korol,S., Liebl,S., Logghe,M.,
 Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Maillet,L.,
 Manhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S.,
 Nasr,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E.,
 Pohl,F.M., Pohl,T.M., Potier,S., Proft,M., Purnelle,B., Ramezani
 Rad,M., Rieger,M., Rose,M., Schaffl-Gerstenschlaeger,I.,

Schierens, B., Schwarzlose, C., Stela, J., Slonimski, P. P., Smits, H. M., Souciet, J. L., Steensma, H. Y., Stucka, R., Urrestalaz, A., van der Aalst, Q. J., van Dyck, L., Vassarotti, A., Vetter, I., Viendendaele, F., Vissers, S., Wagner, G., de Wierflose, P., Wolter, K. H., Zegulski, M., Zimmermann, F. K., Mewes, H. W. and Kleene, K. Complete DNA sequence of yeast chromosome II
EMBO J. 13 (24), 5795-5809 (1994)
95112786

source

1. .1688

Pohl, F. M., Pohl,

T.M., Potier, S., Profit

REFERENCE 1 (bases 1 to 3709)
 AUTHORS Benoit, B., Nemeth, A., Auner, N., Kuhn, U., Simonelig, M., Wahle, E., and Bourbon, H. M.
 TITLE The Drosophila poly(A)-binding protein II is ubiquitous throughout Drosophila development and has the same function in mRNA polyadenylation as its bovine homolog in vitro
 JOURNAL Nucleic Acids Res. 27 (19), 3771-3778 (1999)
 MEDLINE 99412419
 REFERENCE 2 (bases 1 to 3709)
 AUTHORS Bourbon, H. M.
 TITLE Direct Submission
 JOURNAL Submitted (23-DEC-1998) UMR5540 du CNRS, CNRS, 118 Route de Narbonne, Toulouse 31062, France
 FEATURES
 source
 1. 3709
 /organism="Drosophila melanogaster"
 /strain="Oregon R"
 /db_xref="taxon:7227"
 /chromosome="2"
 /map_2R: 448"
 join(257..562,761..972,1228..1629,1780..1913,1979..2046,2472..3226)
 /gene="Pabp2"
 /product="poly(A)-binding protein II"
 257..3226
 /gene="Pabp2"
 join(874..1629,1780..1913,1979..2018)
 /gene="Pabp2"
 /function="RNA-binding protein"
 /codon_start=1
 /product="poly(A)-binding protein II"
 /protein_id="AA00976.1"
 /db_xref="GI:6007612"
 /translation="MADEDITLNEDELLEETNGEOTEIATEVEEGSMOIDPEL
 EAKKAKKEEKEEAEKIKOMOSEVDKATVPLSLSEKQEDTRESVYGN
 VYGAABEIEBAFHGCGTINRTITCNKADGPKFATIERESKPEVETALAMETL
 FRQRIKVMKRTNRPGLSTTNRFANGSFGARVSRACSHSTFGARRAMKVRGA
 NYAPY"
 BASE COUNT 1146 a 777 c 785 g 1001 t
 ORIGIN
 Query Match 0.78; Score 20; DB 3; Length 3709;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2499 tgcacatcgatgaaatcc 2518
 ||||||||||||||||||
 Db 684 TGCACATCGACTGAAATCC 665
 RESULT 55
 SCIRAL/C
 LOCUS SCIRAL 12595 bp DNA PLN 22-SEP-1994
 DEFINITION S.cerevisiae (S288C) IRAL1, YBR1118 and YBR1119 genes.
 ACCESSION X78937
 VERSION X78937.1 GI:547575
 KEYWORDS DEAD box protein; helicase; IRAL gene; YBR1118 gene; YBR1119 gene.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 12595)
 Zaguinski, M., Becam, A. M., Grzybowska, E., Lacroute, F., Migdalski, A.,
 Slonimski, P. P., Sokolowska, B. and Herbert, C. J.
 TITLE The sequence of 12.5 kb from the right arm of chromosome II predicts a new N-terminal sequence for the IRAL protein and reveals two new genes, one of which is a DEAD-box helicase
 JOURNAL Yeast 10 (9), 1227-1234 (1994)
 MEDLINE 95274325
 REFERENCE 2 (bases 1 to 12595)
 AUTHORS Herbert, C. J.
 TITLE Direct Submission

JOURNAL Submitted (21-APR-1994) C.J. Herbert, Centre de Genetique Moleculaire, CNRS, 91198 Gif-sur-Yvette, FRANCE
 FEATURES
 source
 1. 12595
 /organism="Saccharomyces cerevisiae"
 /strain="S288C"
 /db_xref="taxon:4932"
 /chromosome="II"
 /clone_lib="cosmid"
 complement(1..8331)
 /gene="IRAL"
 complement(<1..8331)
 /gene="IRAL"
 /codon_start=1
 /protein_id="CAA5537.1"
 /db_xref="GI:547576"
 /db_xref="SWISS-PROT:P18963"
 /translation="MNSDPQDKKPNMYSITKHLFPRRLLVLPESNLTAYDVE
 ADSVENSRSILNLTATKDLNPIENTGLIDLVODEITSDNTDIAISLVLL
 RLSDVEFYWDQNDPKIRNDNKPQSSRPNPHSPKPTRINPALAMVLCKI
 SKLENTRTILKVLONMSHLSCSATISKSSILPDSQEFLOKKNYAYTEKIDLTIDYI
 OREISASNHVEFTKCVKRVVAPLISHTSTELGVYVNHDLPGCEYLTKGLAYIDI
 LQHSYMKRTIEFHSILYYASKAPLFMTMARPKYVKYNNLISDVSPPSSSDGN
 GSNSDKTISISLVSLFDVYSTVSSSLINVNDHHLHHSSSSKTTNTSPN
 SISKTSIKOSVNAAGNVSPSTGNDASPTSPASLSPPLNTNLGLPISPTSTL
 GOANSTSTTAATTKTDADPTMTNTNNNNNSANLNINIPORTIFSLDDISFNSR
 KSLINDSNSLFMTDSOHSNMTNNAHACVNSOSQNDOSILNEMELYSN
 TSELSHHTAILRELVVLFLDSEYDEKNSYRKISEPININDSPDSMSAS
 KPSIRHLTHGLKLTLOOGRKRNKVELYILRNGQFVDGLSIDSRILMT
 MTSISQIDSNIAVSIFSKRYNLQONLEVTNMSATANTFISHCERNPLTHRL
 QLEFFASGLQSDLELRLHOLEKELNHLDEKISLYTEGFVFFHLVSTKLHSDIA
 EKSSVLRKLFICIIDILKATPEPDNDTKIISLIDHIDPDARTLSTKLHSDIA
 FPAATSYTEPEIEIINSSDASVLSOSPISNGSNITWTMTDQSIPTLSNR
 SSASDLSLNTILNPLEAONNANLHLRSGVPTTKRASPNDSEKRSQSPSP
 QLOQSDPLSPVSSSAGFSSNHTATPTLTKIKSPKTKKIDAKLOKPSY
 SRYTISDNDKARKMNTSTIRKRTNMTIRPDANTEPKTTIDIKPLPVSITLSN
 RLQVVARARTEIPLSTAFEDINDLDNRVANDHLYLCTVAVTLPASSLFOLKLENA
 KREMLDIIIVKQVRYSYLSNAEKHNLQVATITTERLPLPVAAGSGIPIISYCS
 RGNTPLRIKISCEFLRSFYOKYVGAIDQYSIYIDFIDMAQDNFASGVALOR
 RLNNILITYIKGSDSILDSMDVIYKKMFYFSGSKSVQDEELVDFRSAGIILASNGI
 LSDMOLEKSKAPONEGDSLSPESKNPAYEVKSLKELTKMNFISKOCQNNP
 NLITRNSDILSTELHPLSNPLNPLNLSKIDELSDLSKSHEDSVLVLPOTIIT
 IRTTLKRDDERKIMLFSTDLDAVKKLEIYEKISIKSKYKGILOMSKRFKEH
 SEKNLGISNHPHKKNMKLVIGWELSTNKYDENVSRPLRENDLOKRDDEPLYID
 TSIESAKALAYLTHNVLEIIPSSKEDMNRSSYFGFHPFILLKGLKESADLNOFP
 VSLRHKISILENVIYALTNLSNANYVNLKFTLPBGYPSPNDIRIAFTEIDVTN
 YVNPPEKHEMDKMLADDFLYIKRNPILAFPGSLACSPADVDVLAGFLNMPDRNA
 SHLYTEILKOEIKRARSDILRRSCATRALSYTRRGKKYILKTRAPLOGIVD
 NKESEFIDKMPGENSEKMDLFEKRYMRLIDATSSIDDPFIEVDICKITIVNAS
 VNEPEYAYIAVGSFVLRFIGPALVPSDENIIIVHADRPFTILAKVIOSLANGR
 ENIFKKDILVSKKEEFLKCSDKIFNLSLCKIPNNFVNVREDDPTISPDSFLHK
 FFLYNEFTIRKETINESKLGEESFLKNTVMNDKILGLVQSPMEIKNEIPYVEN
 RKQYPSLYEFMSRYARFKVDMKEEDNAPVHMAVTLDGIIIVTFEINCYNFMV
 DELVYVLDIYIARMCSSKRYVIDCTTFPGGKANQKLTTFEFLIPBASNCAGCY
 YFNVNSFMDQNASSTYVENPVLVTPPCFINSNDOSILSLGSGSLGSLADIVR
 VTLHDITLYDKKKKCPVSLKIGNKYFQVILHIDPLXVYVSNTFISKFNVYKIS
 NLISVDVNTGVSSEFLTNEEKELVCSKRYEIVMFYADLKMEEDPDEFSN
 DISFSSSNAVNASYCNKVEGEIISHLSLVGLENDFLVKNISVNLAVATGEP
 NIDFGRLKSPETVDPDTTFILALIFAFSSSESLTPYIKMKLLENDVYPOE
 HIPTVYCSLSTVNPVTEHYVLANDEGEALSRITISRLTYKERNPTTYALDOIV
 FLTLADGRLLNVIETIVSHALDRSDENDKAAVILSPFTIACQVETKLINMT
 KSLPSLAEVAASHVSELTLTILSKISVSIFPSPPLSQVYDEILFASLLIDVQSP
 IRVSLYTEILMNVCHITNESLPERNRKMLDVCATFAROKLNFISGSEQGRYALP
 PAASSSSKFTGLDFTKNIIMLEIGSISEQADWAKKVKYLMDAIFGRHSFESARA
 NMIGIMSKSHSLDFLCKRLVETKVFAPVVDDEQMIIITAHFTYSIKVIGLSDPS
 SELMKEL"
 complement(8728..9741)
 /gene="YBR1118"
 complement(8728..9741)
 /gene="YBR1118"
 /codon_start=1
 /protein_id="CAA5538.1"
 /db_xref="GI:547577"

once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone R01E6 is at 1 in this sequence. The true right end of clone R01E6 is at 340 in sequence Z70208.
The true left end of clone F54B11 is at 40834 in this sequence. The true right end of clone F18H3 is at 23326 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z50110.
The end of this sequence (40834..40937) overlaps with the start of sequence Z70208.
for a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=R01E6>.

FEATURES

source

gene

CDS

```
1..40937
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="R01E6"
join(999..1127,1467..1589,1647..1756,1803..1905,1958..2060,2118..2307,2362..2503,2557..2747,3149..3251,3304..3441,3489..3656,4047..4244)
/gene="R01E6.4"
join(999..1127,1467..1589,1647..1756,1803..1905,1958..2060,2118..2307,2362..2503,2557..2747,3149..3251,3304..3441,3489..3656,4047..4244)
/gene="R01E6.4"
```

```
/note="contains similarity to Pfam domain: PF00065
(Neurotransmitter-gated ion-channel), Score=484.8,
E-value=2.2e-142, N=1
```

```
cdna EST yk268h8.3 comes from this gene
cdna EST yk268h8.5 comes from this gene
cdna EST yk354e9.3 comes from this gene
cdna EST yk354e9.5 comes from this gene
```

```
/codon_start=1
```

```
/protein_id="CA92184.1"
```

```
/db_xref="GI:3878769"
```

```
/translation="MSFNMALLILVAVGSTNAFKIKRAKDLESQIYEDLLDYKNK
```

```
PRPVKNSDILTPDVGASLIRIIDVDKNOVLTTNLMLEKMNDAKLTWPEKYGIGK
TLHIPDFIWTPLVLYNNAAGPDITILDALVTEGNYWQPPAIYKSCFPIDYVM
PYDSQCEKMEKFTWYTGKRYVDLKLPOEEVYITKDDNDFEQMOGDLSPFYRSA
EMDLSTSEHSVLAASCGGPKYVDITVEYFGIRKRTLFETCNLLPCPLISLTTF
VEYLSHKITFESISLWTLVTFVLVLDLMPRTSVIPMGRLLITMIIIVASTVYS
VITVNRERNGSAHKSMPWIRAVLFKFLKILMSREKGEVTKQPLVLAASLVGSN
YATPAKAANEYRNKAKGKNDLMSLRQSPQAVLNTFDMRMNRNDSTVENLKPVL
CAONNNAPTVAAPRTKROKGSYDDVFMNLNQVRFIAEHFRHNELEGISDWTFFV
AMVLDRFLIFIVLVNGVFIILLESPLDYSCPMNITVFNKPLGOANFYSSMNLK"
```

```
join(5029..5054,5131..5546,5594..5889)
```

```
/gene="R01E6.5"
```

```
join(5029..5054,5131..5546,5594..5889)
```

```
/note="cdna EST yk146b3.5 comes from this gene
```

```
/gene="R01E6.5"
cdna EST yk170c5.5 comes from this gene
cdna EST yk146b3.3 comes from this gene
cdna EST yk170c5.3 comes from this gene
cdna EST yk412b8.3 comes from this gene
cdna EST yk412b8.5 comes from this gene
cdna EST yk387b5.3 comes from this gene
cdna EST yk422c1.3 comes from this gene
cdna EST yk422c1.5 comes from this gene
cdna EST yk309d12.3 comes from this gene
cdna EST yk309d12.5 comes from this gene
```

```
/codon_start=1
```

```
/protein_id="CA92185.1"
```

```
/db_xref="GI:3878770"
```

```
/translation="NASPISHSHENASGKYGDSQHGKNFHOKAEHSGHAKKRG
```

```
YSODGKAADHDALHKDAAEKOKKEAHQSSKRYTDDASNTYKKDYTKTFFGFDTRY
VQOYVMEQYHTDEKHNANKYAGDEHNAQOQHKNDGSHADGOGYDKADKHAHSHTN
YGNEDAGHKSQYDNEGKYGDHTDGYNNKGYDNYGOGHDESEYQSPSYQPKPKRYGCH
OSHYPQARSHPYQGDQYGYEAPHY"
```

gene

CDS

```
join(11014..11029,11105..11235,11321..11420,11472..11686,11940..12167,12215..12304)
/gene="R01E6.3"
join(11014..11029,11105..11235,11321..11420,11472..11686,11940..12167,12215..12304)
/gene="R01E6.3"
/note="contains similarity to Pfam domain: PF00194
(Eukaryotic-type carbonic anhydrase), Score=144.1,
E-value=8e-40, N=1
```

```
cdna EST yk202f1.3 comes from this gene
```

```
cdna EST yk202f1.5 comes from this gene
```

```
cdna EST yk234h4.3 comes from this gene
```

```
cdna EST yk234h4.5 comes from this gene
```

```
/codon_start=1
```

```
/protein_id="CA92190.1"
```

```
/db_xref="GI:3878775"
```

```
/translation="MPRKERYSKTQSSRNERLKTFTPLAAQROSPIDIVPOHCCD
```

```
TDVCKADALNIDYKSGDCDPLVSEGGFVNVNRCNGTFLANHLPSKFKALQAFH
WGSNSKSGSEHFLDQKOLSGEVHFEVWNTSYESFNVALSKDGLAVGVKRGKYN
NYHGLIDYVRKATGNATPRTAMPDPHEHILSPDKREFYTGSLTPPYNCCVWT
LTFEPVSVSGQNLVLRNITPANHRCQDKCEIKSSNF"
complement(join(13219..13506,13552..13698,13747..13815,13862..13985,14486..14626,14675..14843,14895..15086,15323..15497,15549..15620,15675..15963,16018..16099,16292..16364,16408..16501,16800..16951,17049..17172,17430..17501,17549..17721,17789..17885,17932..18006,18054..18420,18589..18656))
/gene="R01E6.1"
```

CDS

gene

```
complement(join(13219..13506,13552..13698,13747..13815,13862..13985,14486..14626,14675..14843,14895..15086,15323..15497,15549..15620,15675..15963,16018..16099,16292..16364,16408..16501,16800..16951,17049..17172,17430..17501,17549..17721,17789..17885,17932..18006,18054..18420,18589..18656))
/gene="R01E6.1"
/note="contains similarity to Pfam domain: PF00069
(Eukaryotic protein kinase domain), Score=25.7,
E-value=1.6e-06, N=2; PF00211 (Adenylate and Guanylate
cyclase catalytic domain), Score=191.2, E-value=5.5e-54,
N=1"
```

```
/codon_start=1
```

```
/protein_id="CA92186.2"
```

```
/db_xref="GI:14530534"
```

```
/translation="WCVLRMQMRPRPDSVOYSSEKHELLTVIILKFCGVHRI
```

```
NOHQSGDQNSVSAASYAINAVAFVPGTCTDTRITGDAEIKRSEVI
GYPRVEARQVQELTVINVAQFSGVAGETLVFLKKELEQVETLVGSAVKVPLNLS
LSNDLSYVNIIMSEFKIREVEVDENDVMTYKQDKIKGAARMIVVCAQPDYISAFY
NIGIRBELSGREFITVILINKPDEILNORVKRLKLSNAFTISPIQEOYSDAFSIMO
DYTPNLADQFTFFELNITHACTAYCVGSVNGAETQTDNHTASGKAVTTKGTFTFD
NSGSVLNTNAVFITINPAEMTFESIILTLKSANSQCDYNEFQTSNPTSLDMLTKMD
PPDDCAKSSCVNYIPHIAAVIVITVIAIVIVKORRHKNITKWKYKESLK
IIVKNADNAKQRELENRASNTDNAALTSRPFVSFVQIDLOQVAIRFADLADQMTMR
FPEFTTLDYLSIKOLOHNDLAKYGIQVNDWITMIIIDHVLERGTLIEFCIDRDQMG
DQFEKAPMRDILKGLQVILKSSIGYHGLQWSTCGLIDHMLKTLTVXSNPMNSQL
DAENITVPEQAAHMITTPQVYCPPEPHIRETDSGQPPRYVRGSPKGIITYCGKRTF
```

```
MYVEREDPYHLIHSVERPNATLILKQILNEHSEKRAITDDYRQENMLLEWKECDMDRP
KRPITIKLEISTVYPLSKGNLVQDMIRKEMKAPIDELQOMVAIRFADLADQMTMR
LNLNEMLPASIAKDLKNGLIIMPRESYSAVMEFQIDFNALMKRSSREPOVIAFLINDY
DOFDYIKRHDAAKVEGTGTETVAVASGVHENGRIIFVAVESLEISREIYVIAVOH
DKMYKIRIRIGPFIAGPAGIAGVGIGIRSPRCICDPTVAFSPRMSQNCPPNOIQTSETTA
```

```
RLLPDSHEKFKYVRGIVYHVGKGNNAKRLKICCTPFETHSIDL"
```

```
complement(join(20486..20551,21018..21101,21336..21491,21729..21881,22959..23081,23639..23753,23802..23950,23999..24088))
```

```
/gene="R01E6.6"
```

```
complement(join(20486..20551,21018..21101,21336..21491,21729..21881,22959..23081,23639..23753,23802..23950,23999..24088))
```

```
/gene="R01E6.6"
```

```
/note="similar to 951004: R01E6.6 has homology with
```

```
C36E8.2"
```

```
complement(join(20486..20551,21018..21101,21336..21491,21729..21881,22959..23081,23639..23753,23802..23950,23999..24088))
```

```
/gene="R01E6.6"
```

```
/note="similar to 951004: R01E6.6 has homology with
```

```
C36E8.2"
```

```
complement(join(20486..20551,21018..21101,21336..21491,21729..21881,22959..23081,23639..23753,23802..23950,23999..24088))
```

```
/gene="R01E6.6"
```

```
/note="similar to 951004: R01E6.6 has homology with
```

```
C36E8.2"
```

```
complement(join(20486..20551,21018..21101,21336..21491,21729..21881,22959..23081,23639..23753,23802..23950,23999..24088))
```

```
/gene="R01E6.6"
```

```
/note="similar to 951004: R01E6.6 has homology with
```

```
C36E8.2"
```

```

/codon_start=1
/protein_id="CAA92187.1"
/db_xref="GI:3878772"
/translation="MGNSSSSSRIRKKSESMETREKPNRSSVGFENKSKGMDMSS
KSRSTSTASNTSRKSTSTTKDKDRPKSSSEKNIETSPRKSNKDKDHPISAGRE
ITTCCTENHSEFANKVYORITEKREDYKTYINLAKERSIVNNRLKOLVEDIVAH
HDADPLESVKQYGEHVELKQYGEFDPWVAADAMLEGVILDMANHPADTVSAW
SLSLVMTFSSVDGYSELRRHMSRRLKQSTYDSRENNEDTPEGIHSEVCCHI
PLSPHHPANENVLTLLORLLQAAQPDARSSNNL"
complement(join(33511..33756,33804..33947,33995..34096,
34158..34242,34362..34369))
/gene="R01E6.2"
complement(join(33511..33756,33804..33947,33995..34096,
34158..34242,34362..34369))
/gene="R01E6.2"
/codon_start=1
/protein_id="CAA92188.1"
/db_xref="GI:3878773"
/db_xref="SPTREMBL:Q21619"

```

Query Match 0.7%: Score 20; DB 3; Length 40937;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2379 gttgtattcattgaccc 2398
 |||||
 Db 22674 GTGTTATTTCATTGACC 22693

```

RESULT 58
AC006103 45459 bp DNA HTG 04-DEC-1998
LOCUS Homo sapiens chromosome 10 clone LA10NC01.124_D_3 map 10q25.1, ***
DEFINITION SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION AC006103
VERSION AC006103.1 GI:3962484
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 45459)
AUTHORS Smith,D.R.
TITLE Sequencing of Human Chromosome 10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 45459)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
COMMENT Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
Technology. Data may contain low quality sequence and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 45459: contig of 45459 bp in length.
Location/Qualifiers
1.45459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LA10NC01.124_D_3"
/chromosome="10"
/map="10q25.1"

```

FEATURES

```

source
BASE COUNT 13350 a 9755 c 9636 g 12671 t 47 others
ORIGIN

```

```

Query Match 0.7%: Score 20; DB 2; Length 45459;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2879 tattgaagatcttaagcaa 2898
|||||
Db 8207 TATTGAGATCTTAAGCAA 8226

```

```

RESULT 59
AC005210 83969 bp DNA PRI 05-NOV-1999
LOCUS citb_179_n_3, complete sequence.
DEFINITION AC005210
ACCESSION AC005210
VERSION AC005210.3 GI:6249673
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Sequencing of Human Chromosome 10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
REFERENCE 3 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
REFERENCE 4 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
REFERENCE 5 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
REMARK Vector Sequence Clipped
COMMENT On Nov 5, 1999 this sequence version replaced gi:4314329.
FEATURES
source
1.83969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="CIT987SK-1179N3"

```

```

BASE COUNT 24380 a 19036 c 18255 g 22298 t
ORIGIN

```

```

Query Match 0.7%: Score 20; DB 9; Length 83969;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2879 tattgaagatcttaagcaa 2898
|||||
Db 68688 TATTGAGATCTTAAGCAA 68707

```

```

RESULT 60
AL159169 86155 bp DNA PRI 28-AUG-2000
LOCUS AL159169/c
DEFINITION Human DNA sequence from clone RP11-408A13 on chromosome 9, complete
sequence.
ACCESSION AL159169

```

VERSION AL159169.14 GI:9944142
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 86155)
AUTHORS Collier,R.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 29, 2000 this sequence version replaced gi:9908927.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-408A13 is from the library RPCT-11.2 constructed at the
Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-408A13. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-408A13 is at 86155 in this
sequence. The true left end of clone RP11-439M12 is at 95 in this
sequence. The true right end of clone RP11-1120J1 is at 100 in this
sequence.

FEATURES
source Location/Qualifiers
1..86155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-408A13"
/clone_lib="RPCT-11.2"
complement(1..92)
/note="match: GSS: Em:AQ351225"
104..499
/note="match: GSS: Em:AQ830380"
113..440
/note="match: GSS: Em:AQ720173"
620..831
/note="106 copies 2 mer tt 55% conserved"
921..1076
/note="MIR repeat: matches 2..168 of consensus"
1112..1245
/note="LIM4 repeat: matches 4699..4836 of consensus"
1511..2117
/note="match: GSS: Em:AQ427218"
2745..3053
/note="AluSg repeat: matches 1..309 of consensus"
3152..3341
/note="MIR repeat: matches 48..247 of consensus"
3697..3752
/note="28 copies 2 mer ac 85% conserved"
4146..4327

repeat_region /note="MERSB repeat: matches 7..178 of consensus"
4708..4847
/note="MIR repeat: matches 92..241 of consensus"
4899..5065
/note="MERSA repeat: matches 3..189 of consensus"
5103..5207
/note="MERSA repeat: matches 65..166 of consensus"
5132..5233
/note="MERSA repeat: matches 9..112 of consensus"
7190..7282
/note="L2 repeat: matches 2331..2421 of consensus"
7384..7439
/note="L2 repeat: matches 2577..2634 of consensus"
9060..9293
/note="MIR repeat: matches 3..243 of consensus"
9777..10318
/note="match: GSS: Em:AQ395702"
complement(10404..10831)
/note="match: GSS: Em:AQ30903"
10804..10990
/note="MIR repeat: matches 47..231 of consensus"
14497..14804
/note="match: GSS: Em:AQ665295"
14510..15126
/note="match: GSS: Em:AQ310218"
14782..14976
/note="MIR repeat: matches 66..256 of consensus"
15854..16584
/note="match: GSS: Em:AQ394111"
17012..17077
/note="L2 repeat: matches 2680..2745 of consensus"
19799..19828
/note="15 copies 2 mer ga 93% conserved"
20263..20342
/note="40 copies 2 mer aa 68% conserved"
23323..23412
/note="L2 repeat: matches 2594..2688 of consensus"
24826..24898
/note="MIR repeat: matches 46..119 of consensus"
24991..25058
/note="34 copies 2 mer ga 67% conserved"
25396..25600
/note="MIR repeat: matches 1..217 of consensus"
26063..26184
/note="L2 repeat: matches 2572..2710 of consensus"
27687..27775
/note="L1MD3 repeat: matches 7659..7739 of consensus"
28585..28788
/note="L2 repeat: matches 2558..2749 of consensus"
28834..29034
/note="AluY repeat: matches 1..201 of consensus"
29035..30250
/note="LIM4 repeat: matches 3018..4252 of consensus"
29792..30537
/note="match: GSS: Em:AQ353410"
29795..30640
/note="match: GSS: Em:AQ898185"
30938..31350
/note="match: GSS: Em:AQ627562"
32743..32817
/note="MIR repeat: matches 187..262 of consensus"
33146..33212
/note="L2 repeat: matches 2679..2750 of consensus"
35009..35139
/note="LIMAS4 repeat: matches 6162..6294 of consensus"
35486..35885
/note="MSRB repeat: matches 1..426 of consensus"
35886..37460
/note="MSRB-internal repeat: matches 1..1651 of consensus"
37463..37675
/note="MSRB repeat: matches 203..420 of consensus"
37685..37913
/note="MSRB repeat: matches 1..229 of consensus"

```

repeat_region      39397..39444
                    /note="24 copies 2 mer tl 75% conserved"
repeat_region      40066..40259
                    /note="LIM4 repeat: matches 4385..4593 of consensus"
repeat_region      40534..40833
                    /note="LIM3 repeat: matches 5866..6180 of consensus"
repeat_region      41355..41394
                    /note="20 copies 2 mer aa 82% conserved"
repeat_region      42168..42231
                    /note="L2 repeat: matches 2642..2705 of consensus"
repeat_region      43215..43505
                    /note="Aluub repeat: matches 1..307 of consensus"
repeat_region      43933..43969
                    /note="L2 repeat: matches 2696..2732 of consensus"
repeat_region      43935..43986
                    /note="MIR repeat: matches 216..260 of consensus"
repeat_region      44879..45241
                    /note="L2 repeat: matches 2264..2709 of consensus"
repeat_region      45313..45472
                    /note="MIR repeat: matches 4..183 of consensus"
repeat_region      45633..45946
                    /note="Aluub repeat: matches 3..312 of consensus"
repeat_region      46261..46675
                    /note="HAL1 repeat: matches 1098..1552 of consensus"
repeat_region      46676..46713
                    /note="L9 copies 2 mer aa 81% conserved"
repeat_region      46993..47895
                    /note="L2 repeat: matches 1181..2181 of consensus"
repeat_region      48487..48532
                    /note="23 copies 2 mer ta 76% conserved"
repeat_region      48581..48646
                    /note="33 copies 2 mer at 68% conserved"
repeat_region      48824..49187
                    /note="THE1C repeat: matches 1..371 of consensus"
repeat_region      49538..49616
                    /note="MER5B repeat: matches 275..336 of consensus"
repeat_region      49617..49972
                    /note="LIMB2 repeat: matches 5810..6163 of consensus"
repeat_region      50230..50439
                    /note="match: GSS: Em:B88855"
repeat_region      50459..50934
                    /note="match: GSS: Em:AQ247716"
repeat_region      51761..52062
                    /note="MER3 repeat: matches 3..323 of consensus"
repeat_region      54014..54039
                    /note="13 copies 2 mer ca 100% conserved"
repeat_region      54903..54981
                    /note="MIR repeat: matches 51..139 of consensus"

Query Match
Best Local Similarity 0.7%; Score 20; DB 9; Length 86155;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1927 tacattccagcagggctgt 1946
      |||||
Db 62606 TACATTCACGACGAGCTGT 62587

```

```

AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Homo sapiens 86,719 genomic DNA of 21q22.1
JOURNAL      Published Only in Database (1999) In press
REFERENCE    2 (bases 1 to 86719)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (15-DEC-1999) to the DDBJ/EMBL/GenBank databases.
              Masahira Hattori, The Institute of Physical and Chemical Research
              (RIKEN), Genomic Sciences Center (GSC): Kitasato Univ., 1-15-1
              Kitasato, Sagamihara, Kanagawa 228-8555, Japan
              (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-42-778-9923, Fax:81-42-778-9924)
FEATURES     Location/Qualifiers
              source          1..86719
                              /organism="Homo sapiens"
                              /db_xref="taxon:9606"
                              /chromosome="21"
                              /clone="B680H4"
                              /map="21q22.1"

BASE COUNT   26401 a 16034 c 16025 g 28259 t
ORIGIN

```

```

Query Match
Best Local Similarity 0.7%; Score 20; DB 9; Length 86719;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2430 catatgcagtcctctgtcaa 2449
      |||||
Db 3109 CATAATGCCAGCTCTGTCAA 3128

RESULT 62
LOCUS      AC005923      88326 bp      DNA      PRI      12-JAN-2000
DEFINITION Homo sapiens 3p21.3-4 PAC RP4-751E10 (Roswell Park Cancer Institute
              Human PAC Library) complete sequence.
ACCESSION  AC005923
VERSION     AC005923.2 GI:4309927
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 88326)
REFERENCE   1 (bases 1 to 88326)
AUTHORS     Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,J., Chen,Z.,
              Culpepper,P., Ding,Y., Dugan,S.P., Durbin,K.J., Forcum,J.,
              Ganesh,R.P., Garcia,C., Garcia,D.K., Gorrell,H., Gorrell,L.L.,
              He,X., Hernandez,J., Jackson,L.E., Kondajewski,N., Leal,B.,
              Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Moore,S.,
              Moorish,T., Nguyen,N., Oswal,G., Pamell,L.R., Parish,B.J.,
              Perez,L.M., Rashid,N.D., Rives,C.M., Scherer,S.E., Shen,H.,
              Simon,H.L., Vo,Q.K., Wei,Y., Williamson,A.L., Worley,K., Zhou,X.,
              Naylor,S.L. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 88326)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-1998) Molecular and Human Genetics, Baylor
              College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
              3 (bases 1 to 88326)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (27-FEB-1999) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              4 (bases 1 to 88326)
REFERENCE   4 (bases 1 to 88326)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (12-MAR-1999) Human Genome Sequencing Center, Department

```



```

repeat_region      /rpt_family="Alusx"
                   complement(47189..47498)
repeat_region      /rpt_family="AluY"
                   complement(47499..47782)
misc_feature        47716..47800
                   /note="Sequence is generated from PCR product only"
                   /function="Low coverage"
misc_feature        47802..48101
                   /function="Gap of approximately 300 bps"
repeat_region      complement(48158..48460)
gene               /rpt_family="Aluub"
                   join(48873..49019,49455..49645)
                   /gene="Unigene cluster containing AA292850 and AA405356"
repeat_region      49867..50178
                   /rpt_family="Aluub"
repeat_region      complement(50948..51140)

Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 88326;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1926 ctacattccagcagcgctg 1945
Db 72992 CTACATTCAGCAGCGCTG 73011

RESULT 63
LOCUS AC084447 88839 bp DNA INV 04-NOV-2000
DEFINITION Caenorhabditis briggsae cosmid CB019G12, complete sequence.
ACCESSION AC084447
VERSION AC084447.1 GI:11094897
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae.
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditiidae; Rhabditiidae; Pelodermidae; Caenorhabditis.
REFERENCE 1 (bases 1 to 88839)
          Washington University Genome Sequencing Center.
          The C. briggsae Genome Sequencing Project
          Unpublished
          2 (bases 1 to 88839)
          Waterston,R.
          Direct Submission
          Submitted (04-NOV-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
          Genome Sequencing Center
          Department of Genetics, Washington University,
          St. Louis, MO 63110, USA
          e-mail: jsplethwatson.wustl.edu

```

FEATURES

```

SOURCE
1..88839
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="CB019G12"
9330..9401
/note="codon recognized: CCA"
trna 9814..9885
/product="tRNA-Pro"
trna 30927..30998
/product="tRNA-Pro"
/note="codon recognized: GGA"
/product="tRNA-Gly"

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

```

BASE COUNT 28080 a 16056 c 16645 g 28058 t
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Score 20; DB 3; Length 88839;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2714 aggtccacacgcgcgaag 2733
Db 67631 AGGTCCACACGCTGTCAG 67612

```

```

RESULT 64
LOCUS AL391278 91733 bp DNA HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-885P2, *** SEQUENCING IN
          PROGRESS ***, 29 unordered pieces.
ACCESSION AL391278
VERSION AL391278.10 GI:12331116
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 91733)
          McIay,K.
          Direct Submission
          Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
          On Jan 22, 2001 this sequence version replaced gi:11023650.
          ----- Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquerry@sanger.ac.uk
          ----- Project Information
          Center project name: dj885P2
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid; L08752; 100% of reads
          Chemistry: Dye-terminator Big Dye; 85% of reads
          Dye-terminator Big Dye; 85% of reads
          Consensus quality: 76799 bases at least Q40
          Consensus quality: 81861 bases at least Q30
          Consensus quality: 85494 bases at least Q20
          Insert size: 88933; sum-of-contigs
          Insert size: 169449; 2.6% error; agarose-fp
          Quality coverage: 2.56x in Q20 bases; sum-of-contigs Quality
          coverage: 1.81x in Q20 bases; agarose-fp
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 29 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          1
          2439 2538: contig of 2438 bp in length
          * 2439 2538: gap of 100 bp
          * 2539 4657: contig of 2119 bp in length
          * 4658 4757: gap of 100 bp
          * 4758 7026: contig of 2269 bp in length
          * 7027 7126: gap of 100 bp
          * 7127 9886: contig of 2760 bp in length
          * 9887 9986: gap of 100 bp
          * 9987 12786: contig of 2800 bp in length
          * 12787 12886: gap of 100 bp
          * 12887 15825: contig of 2939 bp in length
          * 15826 15925: gap of 100 bp
          * 15926 17982: contig of 2057 bp in length

```



```
* 17983 18082: gap of 100 bp
* 18083 24394: contig of 6312 bp in length
* 24395 24494: gap of 100 bp
* 24495 27186: contig of 2692 bp in length
* 27187 27286: gap of 100 bp
* 27287 30171: contig of 2885 bp in length
* 30172 30271: gap of 100 bp
* 30272 32949: contig of 2678 bp in length
* 32950 33049: gap of 100 bp
* 33050 36523: contig of 3474 bp in length
* 36524 36623: gap of 100 bp
* 36624 38651: contig of 2028 bp in length
* 38652 38751: gap of 100 bp
* 38752 41483: contig of 2732 bp in length
* 41484 41583: gap of 100 bp
* 41584 44333: contig of 2750 bp in length
* 44334 44433: gap of 100 bp
* 44434 47116: contig of 2683 bp in length
* 47117 47216: gap of 100 bp
* 47217 49339: contig of 2123 bp in length
* 49340 49439: gap of 100 bp
* 49440 54013: contig of 4574 bp in length
* 54014 54113: gap of 100 bp
* 54114 57401: contig of 3288 bp in length
* 57402 57501: gap of 100 bp
* 57502 60611: contig of 3110 bp in length
* 60612 60711: gap of 100 bp
* 60712 63058: contig of 2347 bp in length
* 63059 63158: gap of 100 bp
* 63159 65257: contig of 2099 bp in length
* 65258 65357: gap of 100 bp
* 65358 67754: contig of 2397 bp in length
* 67755 67854: gap of 100 bp
* 67855 72100: contig of 4246 bp in length
* 72101 72200: gap of 100 bp
* 72201 75487: contig of 3287 bp in length
* 75488 75587: gap of 100 bp
* 75588 80699: contig of 5112 bp in length
* 80700 80799: gap of 100 bp
* 80800 84813: contig of 4014 bp in length
* 84814 84913: gap of 100 bp
* 84914 87462: contig of 2549 bp in length
* 87463 87562: gap of 100 bp
* 87563 91733: contig of 4171 bp in length.
Location/Qualifiers
1. 91733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP5-885P2"
/clone_1lb="RCF-5"
1. 2438
/feature="assembly_fragment:00813"
fragment_chain:1"
2539. 4657
/feature="assembly_fragment:00179"
fragment_chain:1"
4758. 7026
/feature="assembly_fragment:00827"
fragment_chain:2"
7127. 9886
/feature="assembly_fragment:00902"
fragment_chain:2"
9987. 12786
/feature="assembly_fragment:00136"
12887. 15825
/feature="assembly_fragment:00196"
15926. 17982
/feature="assembly_fragment:00205"
18083. 24394
/feature="assembly_fragment:00289"
24495. 27186
/feature="assembly_fragment:00290"
```

```
misc_feature 27287. 30171 /note="assembly_fragment:00329"
misc_feature 30272. 32949 /note="assembly_fragment:00334"
misc_feature 33050. 36523 /note="assembly_fragment:00337"
misc_feature 36624. 38651 /note="assembly_fragment:00435"
misc_feature 38752. 41483 /note="assembly_fragment:00437"
misc_feature 41584. 44333 /note="assembly_fragment:00476"
misc_feature 44334. 47116 /note="assembly_fragment:00505"
misc_feature 47217. 49339 /note="assembly_fragment:00589"
misc_feature 49440. 54013 /note="assembly_fragment:00596"
misc_feature 54114. 57401 /note="assembly_fragment:00665"
misc_feature 57502. 60611 /note="assembly_fragment:00677"
misc_feature 60712. 63058 /note="assembly_fragment:00681"
misc_feature 63159. 65257 /note="assembly_fragment:00758"
misc_feature 65358. 67754 /note="assembly_fragment:00805"
misc_feature 67855. 72100 /note="assembly_fragment:00835"
misc_feature 72201. 75487 /note="assembly_fragment:00879"
misc_feature 75588. 80699 /note="assembly_fragment:00912"
misc_feature 80800. 84813 /note="assembly_fragment:00938"
misc_feature 84914. 87462 /note="assembly_fragment:00953"
misc_feature 87563. 91733 /note="assembly_fragment:00807"
vector_side:right"
clone_end:SP6
BASE COUNT 25532 a 18159 c 18530 g 26693 t 2819 others
ORIGIN
```

```
Query Match 0.7%; Score 20; DB 2; Length 91733;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2561 attatcagaaattacctg 2580
DB 88326 ATTTATCAGAAATTAACCTG 88307

RESULT 65
AC073294 LOCUS 93409 bp DNA HTG 26-MAR-2001
DEFINITION Mus musculus clone C77-32119 strain 129 SV, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC073294
VERSION AC073294.1 GI:8493569
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 93409)
AUTHORS DOE Joint Genome Institute.
JOURNAL Sequencing of Mouse
REFERENCE 2 (bases 1 to 93409)
AUTHORS DOE Joint Genome Institute.
```

TITLE Direct Submission
JOURNAL Submitted (13-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 93409: contig of 93409 bp in length.
Location/Qualifiers
1. 93409
/organism="Mus musculus"
/strain="129 SV"
/db_xref="taxon:10090"
/clone="CT7-32119"
/clone_lib="CITDCU7 mouse BAC library"

BASE COUNT 25162 a 21194 c 20677 g 25391 t 985 others

ORIGIN

Query Match 0.7% Score 20; DB 2; Length 93409;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2137 aacattatctctcatggt 2156
|||||
Db 76311 AACATTATCTCTCATGCT 76330

RESULT 66
AC083819/c DNA HTG 06-NOV-2000
LOCUS Mus musculus chromosome 1 clone RP23-285F20, *** SEQUENCING IN
DEFINITION PROGRESS ***; 7 unordered pieces.
AC083819
AC083819.4 GI:10803522
VERSION HTG; HTGS_PHASE1.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 102165)
AUTHORS Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okumori, G., Carroll, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loubege, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mel, G., Moore, S., Morgan, K., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogul, M., Parish, B.,
Perez, L., Relfer, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, C.,
Worley, K. and Gibbs, R.
Direct Submission
Unpublished
JOURNAL 2 (bases 1 to 102165)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

COMMENT Baylor Plaza, Houston, TX 77030, USA
On Oct 16, 2000 this sequence version replaced gi:10801956.
-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
-----Project Information
Center project name: RP23-285F20
Center clone name: MATO
-----Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodypy; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 95179 bases at least Q40
Consensus quality: 98007 bases at least Q30
Consensus quality: 99497 bases at least Q20
Estimated insert size: 100390; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 24502: contig of 24502 bp in length
* 24503 24602: gap of unknown length
* 24603 40809: contig of 16207 bp in length
* 40810 40909: gap of unknown length
* 40910 63390: contig of 22480 bp in length
* 63390 63489: gap of unknown length
* 63490 78501: contig of 15012 bp in length
* 78502 78601: gap of unknown length
* 78602 89787: contig of 11186 bp in length
* 89788 89887: gap of unknown length
* 89888 96681: contig of 6794 bp in length
* 96682 96781: gap of unknown length
* 96782 102165: contig of 5384 bp in length.
Location/Qualifiers
1. 102165
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP23-285F20"

BASE COUNT 30175 a 21419 c 20896 g 29070 t 605 others

ORIGIN

Query Match 0.7% Score 20; DB 2; Length 102165;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2020 aagcaagatatacatatcat 2039
|||||
Db 51969 AACCAAGATATCATATCT 51950

RESULT 67
AL357352 DNA PRI 20-NOV-2000
LOCUS Human DNA sequence from clone RP11-799G19 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL357352
VERSION AL357352.11 GI:11321993
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 106601)
Phillimore,B.

Direct Submission

Submitted (19-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk

COMMENT

On Nov 23, 2000 this sequence version replaced gi:11225761.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP. Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/MGP/Chr6

RP11-799619 is from the library RPC1-11.3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/

VECTOR: pBac3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-799619. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-799619 is at 1 in this sequence. The true left end of clone RP11-532M4 is at 106502 in this sequence.

FEATURES

Location/Qualifiers

1..106601

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP11-799619"

/clone_id="RPC1-11.3"

misc.feature

/note="match: GSS: Em:A0521292"

961..2386

/note="TRIGGER2 repeat: matches 1..1483 of consensus"

2387..2755

/note="THEIC repeat: matches 1..371 of consensus"

2756..4009

/note="TRIGGER2 repeat: matches 1483..2718 of consensus"

4387..4995

/note="LIP3 repeat: matches 768..1393 of consensus"

4995..6352

/note="LIP2 repeat: matches 4779..6163 of consensus"

9015..9156

/note="L2 repeat: matches 2608..2750 of consensus"

9690..9802

/note="MIR repeat: matches 56..172 of consensus"

10378..10458

/note="MIR repeat: matches 71..153 of consensus"

10787..10832

/note="23 copies 2 mer at 78% conserved"

11644..11807

/note="L2 repeat: matches 2581..2749 of consensus"

11817..11984

/note="MIR repeat: matches 5..186 of consensus"

12435..12545

/note="MIR repeat: matches 8..128 of consensus"

repeat_region

/note="18 copies 3 mer tta 88% conserved"

repeat_region

/note="3 copies 15 mer 93% conserved"

repeat_region

/note="AluY repeat: matches 1..286 of consensus"

repeat_region

/note="L2 repeat: matches 2634..2701 of consensus"

repeat_region

/note="19 copies 2 mer ac 81% conserved"

repeat_region

/note="FLAM_A repeat: matches 1..133 of consensus"

repeat_region

/note="L2 repeat: matches 1683..1850 of consensus"

misc.feature

/note="match: GSS: Em:A0790581"

repeat_region

/note="MTR1J repeat: matches 221..337 of consensus"

repeat_region

/note="MTR5A repeat: matches 47..182 of consensus"

repeat_region

/note="MIR repeat: matches 87..245 of consensus"

repeat_region

/note="MTR1F repeat: matches 91..529 of consensus"

repeat_region

/note="AluSq repeat: matches 1..313 of consensus"

repeat_region

/note="L2 repeat: matches 1728..1906 of consensus"

misc.feature

/note="match: GSS: Em:A0783759"

repeat_region

/note="MIR repeat: matches 4..165 of consensus"

repeat_region

/note="LIP3 repeat: matches 6048..6146 of consensus"

repeat_region

/note="LIMC2 repeat: matches 6245..6328 of consensus"

repeat_region

/note="MTR1A1 repeat: matches 1..354 of consensus"

repeat_region

/note="LIMC1 repeat: matches 5357..6250 of consensus"

repeat_region

/note="LIM4 repeat: matches 3807..5147 of consensus"

repeat_region

/note="L2 repeat: matches 2676..2734 of consensus"

misc.feature

/note="match: GSS: Em:A0756546"

misc.feature

/note="match: GSS: Em:A0113333"

misc.feature

/note="match: GSS: Em:A0408008"

misc.feature

/note="match: GSS: Em:A0563645"

misc.feature

/note="match: GSS: Em:A0235395"

misc.feature

/note="match: GSS: Em:A0117470"

misc.feature

/note="match: GSS: Em:A0710247"

repeat_region

/note="LIM4C repeat: matches 1430..1645 of consensus"

misc.feature

/note="match: GSS: Em:A0622979"

misc.feature

/note="match: GSS: Em:A0622990"

repeat_region

/note="LIMC1 repeat: matches 4942..5389 of consensus"

repeat_region

/note="LIMC1 repeat: matches 3969..4776 of consensus"

repeat_region

/note="24 copies 2 mer ta 97% conserved"

```
repeat_region /note="3 copies 18 mer 85% conserved"  
47890..48044  
/note="LIP repeat: matches 3526..3683 of consensus"  
48082..48365  
/note="LIM4 repeat: matches 3095..3390 of consensus"  
repeat_region 49813..50122  
/note="AlusG repeat: matches 1..295 of consensus"  
50653..50930  
/note="L2 repeat: matches 2008..2286 of consensus"  
51027..51334  
/note="AluY repeat: matches 1..308 of consensus"  
51368..51496  
/note="L1PA10 repeat: matches 6029..6156 of consensus"  
repeat_region 51510..51947  
/note="L2 repeat: matches 1292..1752 of consensus"  
52009..52098  
/note="45 copies 2 mer ga 84% conserved"  
repeat_region 52028..52099  
/note="4 copies 18 mer 91% conserved"  
52282..52518  
/note="L1PA1 repeat: matches 5559..5793 of consensus"  
52519..52860  
/note="L1PA7 repeat: matches 5789..6143 of consensus"  
53493..54159  
/note="match: GSS: Em: B93480"  
53511..53921  
/note="match: GSS: Em: A0011862"  
55163..55556  
/note="MER5B repeat: matches 1..403 of consensus"  
55697..56012  
/note="AlusC repeat: matches 1..305 of consensus"  
56584..57061  
/note="match: GSS: Em: A0887852"  
56588..56963  
/note="match: GSS: Em: A0824739"  
56874..56931  
/note="29 copies 2 mer gt 70% conserved"  
57865..58157  
/note="AlusJ repeat: matches 1..293 of consensus"  
58991..59162  
/note="MER5A repeat: matches 6..189 of consensus"  
59736..59927  
/note="MER5B repeat: matches 1..177 of consensus"  
59928..60263  
/note="L2 repeat: matches 144..499 of consensus"  
60293..60745
```

```

Query Match 100.0%; Score 20; DB 9; Length 106601;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 catcacacatcacagaagaga 290
|||||
Db 58795 CATCAGACATCAGAGAGAGA 58814

RESULT 68
LOCUS AL139036/c
DEFINITION AL139036 109810 bp DNA PRI 26-MAY-2001
Human DNA sequence from clone RP11-285H24 on chromosome
13q22.1-22.3, complete sequence.
ACCESSION AL139036
VERSION AL139036.12 GI:14270405
KEYWORDS HYG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 109810)
Leonagorniert, D.
Direct Submission
REFERENCE AUTHORS
TITLE Submitted (26-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL • CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

```

COMMENT

requests: clonerequest@sanger.ac.uk
On May 31, 2013 this sequence version replaced g1:11265984.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT; Tr: TrEMBL, Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>
RP11-285H24 is from the library RP11-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-285H24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-810G21 is at 109711 in this sequence. The true right end of clone RP11-157H4 is at 100 in this sequence.

FEATURES

```

source
1. .109810
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="13"
   /map="q22.1-22.3"
   /clone="RP11-285H24"
   /clone_lib="RPC1-11.1"
   complement(1..98)
   /note="match: GSS: Em:AQ374965"
   205..514
   /note="match: GSS: Em:AQ092767"
   1091..1201
   /note="MT1H repeat: matches 81..189 of consensus"
   repeat_region
   1395..1522
   /note="MT1H repeat: matches 409..510 of consensus"
   2642..3032
   /note="match: GSS: Em:AQ477479"
   2718..3197
   /note="MT1D repeat: matches 1..502 of consensus"
   3752..3940
   /note="7 copies 27 mer 66% conserved"
   3773..4126
   /note="59 copies 6 mer 99999a 54% conserved"
   repeat_region
   3817..4088
   /note="16 copies 17 mer 59% conserved"
   repeat_region
   3818..4076
   /note="7 copies 37 mer 63% conserved"
   repeat_region
   5037..5248
   /note="M1R repeat: matches 12..248 of consensus"
   complement(6151..6886)
   /note="match: GSS: Em:A2519175"
   complement(6237..6883)
   /note="match: GSS: Em:AQ283095"
   complement(6380..6889)
   /note="match: GSS: Em:AQ581403"
   6623..6699
   /note="NA1 repeat: matches 669..746 of consensus"
   repeat_region
   7701..7870

```

```

repeat_region      /note="MIR repeat: matches 60. .220 of consensus"
8648. .8683
/misc_feature      /note="18 copies 2 mer tt 80% conserved"
complement(9052. .9485)
repeat_region      /note="match: GSS: Em:A0620944"
9149. .9432
repeat_region      /note="AluX repeat: matches 1. .284 of consensus"
9905. .10103
repeat_region      /note="MER63A repeat: matches 1. .210 of consensus"
10116. .10217
/misc_feature      /note="L1ME3A repeat: matches 6016. .6116 of consensus"
11190. .11776
/misc_feature      /note="match: GSS: Em:A2374306"
11196. .11815
repeat_region      /note="match: GSS: Em:A2374726"
13390. .13554
repeat_region      /note="MIR repeat: matches 12. .168 of consensus"
15651. .15963
repeat_region      /note="AluY repeat: matches 1. .306 of consensus"
18346. .18991
/misc_feature      /note="match: GSS: Em:A2516011"
18347. .18878
repeat_region      /note="match: GSS: Em:A0384942"
18568. .18864
/misc_feature      /note="Aluub repeat: matches 3. .306 of consensus"
complement(18803. .19250)
repeat_region      /note="match: GSS: Em:A0801630"
20484. .20699
repeat_region      /note="MIR repeat: matches 11. .252 of consensus"
23381. .23476
repeat_region      /note="L1M4 repeat: matches 4013. .4108 of consensus"
23477. .23773
repeat_region      /note="AluY repeat: matches 1. .299 of consensus"
23774. .24043
repeat_region      /note="L1M4 repeat: matches 4107. .4379 of consensus"
24092. .24357
repeat_region      /note="L1ME3 repeat: matches 5861. .6131 of consensus"
complement(24264. .24689)
/misc_feature      /note="match: GSS: Em:A0451631"
24793. .24874
repeat_region      /note="MIR repeat: matches 67. .145 of consensus"
25008. .25518
/misc_feature      /note="match: GSS: Em:A051147"
25611. .25698
repeat_region      /note="MIR repeat: matches 2. .91 of consensus"
26142. .26439
repeat_region      /note="AluX repeat: matches 3. .307 of consensus"
27548. .27704
repeat_region      /note="MER33 repeat: matches 22. .166 of consensus"
27705. .28031
repeat_region      /note="MER2 repeat: matches 4. .345 of consensus"
28032. .28156
repeat_region      /note="MER33 repeat: matches 166. .324 of consensus"
30049. .30177
repeat_region      /note="FLAM_A repeat: matches 1. .133 of consensus"
31105. .31202
repeat_region      /note="MIR repeat: matches 12. .111 of consensus"
31520. .31675
repeat_region      /note="MIR repeat: matches 100. .261 of consensus"
32359. .32560
repeat_region      /note="L1M4 repeat: matches 5318. .5518 of consensus"
32561. .32861
repeat_region      /note="AluJo repeat: matches 1. .296 of consensus"
32862. .33218
repeat_region      /note="L1M4 repeat: matches 5518. .5905 of consensus"
33290. .33388
repeat_region      /note="L1ME3A repeat: matches 6029. .6129 of consensus"
33458. .33489
repeat_region      /note="16 copies 2 mer aa 100% conserved"
33902. .33951
repeat_region      /note="25 copies 2 mer at 74% conserved"
34308. .34780
repeat_region      /note="L1M8 repeat: matches 5785. .6288 of consensus"

```

```

/misc_feature      35285. .35684
repeat_region      /note="match: GSS: Em:A064667"
36808. .37104
repeat_region      /note="AluY repeat: matches 1. .296 of consensus"
37255. .37308
repeat_region      /note="2 copies 27 mer 94% conserved"
37728. .37913
repeat_region      /note="MER63A repeat: matches 1. .210 of consensus"
38654. .38727
repeat_region      /note="2 copies 37 mer 100% conserved"
39086. .39226
repeat_region      /note="MER5B repeat: matches 29. .158 of consensus"
39258. .39424
repeat_region      /note="MER5B repeat: matches 2. .172 of consensus"
39564. .39981
repeat_region      /note="MULTIC repeat: matches 17. .466 of consensus"
44189. .44535
repeat_region      /note="L1MC/D repeat: matches 5436. .5813 of consensus"
45138. .45321
repeat_region      /note="MIR repeat: matches 81. .260 of consensus"
49524. .49727
/misc_feature      /note="MIR repeat: matches 24. .252 of consensus"
49821. .50326
repeat_region      /note="match: GSS: Em:A0169074"
49836. .50027
repeat_region      /note="L1MC/D repeat: matches 5611. .5813 of consensus"
50250. .50567
repeat_region      /note="AluX repeat: matches 1. .312 of consensus"
51288. .51497
repeat_region      /note="MER2 repeat: matches 1. .209 of consensus"
52175. .52231
repeat_region      /note="MIR repeat: matches 15. .74 of consensus"
53123. .53421
repeat_region      /note="AluSg1 repeat: matches 1. .299 of consensus"
54142. .54423
repeat_region      /note="L1MD2 repeat: matches 6011. .6278 of consensus"
54454. .55060
/misc_feature      /note="match: GSS: Em:A0195363"
54535. .54652
repeat_region

```

Query Match 0.7%; Score 20; DB 9; Length 109810;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2791 ttgtgaagaagacctcgaa 2810
|||||
Db 62720 TTGTGGAAGAACCTCTGAA 62701

RESULT 69
AC010902/c AC010902 114169 bp DNA PRI 25-MAR-2001
LOCUS Homo sapiens clone RP11-549J7, complete sequence.
DEFINITION AC010902
ACCESSION AC010902
VERSION AC010902.4 GI:10440867
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (Bases 1 to 114169)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
2 (Bases 1 to 114169)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (Bases 1 to 114169)
AUTHORS Waterston,R.
TITLE Direct Submission

JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 114169)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Sep 30, 2000 this sequence version replaced gi:7631065.
Center project name: H_NH0549J07.
Location/Qualifiers
FEATURES
source 1..114169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-549J7"
BASE COUNT 34392 a 22524 c 23021 g 34232 t
ORIGIN

Query Match 0.7%; Score 20; DB 9; Length 114169;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 aaagcttttcacagaca 279
|||||
Db 74191 AAAGCTTTTTCACAGACA 74172

RESULT 70
AC022124/c 120733 bp DNA PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2075G19, complete sequence.
DEFINITION AC022124
AC022124.6 GI:14993687
HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:1427276.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
Location/Qualifiers
FEATURES
source 1..120733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2075G19"
BASE COUNT 39434 a 20042 c 22104 g 39153 t
ORIGIN

Query Match 0.7%; Score 20; DB 9; Length 120733;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 aagatattgacattattt 384
|||||
Db 61200 AAGATATTGACATTATTTT 61181

RESULT 71
AC008390/c 121931 bp DNA PRI 20-DEC-2000
LOCUS Homo sapiens chromosome 5 clone CTC-235M3, complete sequence.
DEFINITION AC008390
AC008390.7 GI:11908269
HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 121931)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 121931)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 121931)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 20, 2000 this sequence version replaced gi:7711258.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
Location/Qualifiers
FEATURES
source 1..121931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-235M3"
BASE COUNT 39800 a 20269 c 22375 g 39487 t
ORIGIN

Query Match 0.7%; Score 20; DB 9; Length 121931;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 aagatattgacattattt 384
|||||
Db 62415 AAGATATTGACATTATTTT 62396

RESULT 72
AC010072/c 124347 bp DNA PRI 19-NOV-1999
LOCUS Homo sapiens chromosome 14q31 clone CTD-217314 containing TSHR gene, partial cds; and unknown gene, complete sequence.
DEFINITION AC010072
AC010072.5 GI:6453843
HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
COMMENT
TITLE
JOURNAL
AUTHORS
COMMENT
FEATURES
SOURCE
CDS
CDS
CDS
unsure
mRNA
CDS
misc_feature
variation

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 124347)
Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,
James, R., Kaur, A., Madan, A., Owen, M. P., Ratcliffe, A., Shaffer, T.
and Hood, L.
Sequencing of human chromosome 14
Unpublished
2 (bases 1 to 124347)
Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S.,
Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.
Direct Submission
Submitted (11-SEP-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 124347)
Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,
James, R., Kaur, A., Madan, A., Owen, M. P., Ratcliffe, A., Shaffer, T.
and Hood, L.
Direct Submission
Submitted (19-NOV-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
On Nov 19, 1999 this sequence version replaced gi:6114900.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@u.washington.edu
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Big Dye terminators and primers
Assembly program: Phrap; version 0.990399

Location/Qualifiers
1. 124347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q31"
/clone="CDB-217314"
/clone_1id="Caltech human BAC library D"
/note="This clone overlaps RP11-114N19, Accession
AC007262"
complement(join(42575..42673,43714..43786,47934..48025,
52809..52927,53951..54077,62318..62371))
/note="Intron-exon boundaries defined in relation to EST
A1554064. the closest BLASTX similarity is to a
drosophila potassium channel protein"
/codon_start=1
/product="unknown"
/protein_id="AA09033.1"
/db_xref="GI:6453845"
/translation="MLGRYREYNGQAGATEHLKESLEOSIDQRSQRLNRSGRSI
SVTSLSASDLGGTSGELHPPSPKDGDPGICRMRSGRVGVFVQEDDMTOCH
GPHOSLRDLSSEDRIGDDPRELSRSRSDAEKRRLLEETLNLAQKOEYFFVNS
LGKRYKIOCLSSAEVDNVPINWOCGI"
93810..95860
/note="low quality data"
103634..>103846
/note="This is the 5' end of TSHR, found in M73745"
/product="thyroid stimulating hormone receptor"
103677..>103846
/note="thyroid stimulating hormone receptor"
/codon_start=1
/product="TSHR"
/protein_id="AA09032.1"
/db_xref="GI:6453844"
/translation="MRPADLQLVLLDPRDLGAGMCCSSPCECHQEDPRTYCKDI
QRIPSPSTGYTL"
109619..124347
/note="Overlap with RP11-114N19, Accession AC007262"
110216

/note="217314: c: 114N19: t"
/replace="t"
110755
/note="217314: g: 114N19: a"
/replace="a"
111115
/note="217314: c: 114N19: g"
/replace="g"
114198
/note="217314: g: 114N19: a"
/replace="a"
114690
/note="217314: c: 114N19: t"
/replace="t"
115183
/note="217314: t: 114N19: tatat"
/replace="tatat"
115210..115280
/note="low quality data"
116847
/note="217314: t: 114N19: c"
/replace="c"
117131
/note="217314: g: 114N19: a"
/replace="a"
117408
/note="217314: a: 114N19: c"
/replace="c"
117633
/note="217314: g: 114N19: t"
/replace="t"
118901
/note="217314: a: 114N19: g"
/replace="g"
118988
/note="217314: c: 114N19: t"
/replace="t"
119250
/note="217314: a: 114N19: g"
/replace="g"
119718
/note="217314: g: 114N19: t"
/replace="t"
120914
/note="217314: c: 114N19: a"
/replace="a"
121395..121413
/note="217314: c: 114N19: a"
/replace="a"
123019
/note="217314: g: 114N19: a"
/replace="a"
123721
/note="217314: c: 114N19: t"
/replace="t"
123900
/note="217314: t: 114N19: c"
/replace="c"
123926..123927
/note="217314: tt: 114N19: t"
/replace="t"
124015
/note="217314: t: 114N19: c"
/replace="c"
124182
/note="217314: g: 114N19: a"
/replace="a"
BASE COUNT 43520 a 24500 c 22595 g 33732 t
ORIGIN
Query Match 0.7%; Score 20; DB 9; Length 124347;
Best Local Similarity 100.0%; Pred. No. 77;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 atgcacattatctaact 389
 |||
 Db 99037 ATTGACATTAATTTTACTT 99018

RESULT 73
 AC004829
 LOCUS
 DEFINITION Homo sapiens PAC clone RP4-530J23 from 7q21.1-q21.3, complete sequence.
 ACCESSION AC004829
 VERSION AC004829.2 GI:4309818
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Sulston, J.E., and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 REFERENCE
 AUTHORS Harmon, G., Langston, Y., and Kellen, J.
 TITLE The sequence of Homo sapiens PAC clone RP4-530J23
 JOURNAL Unpublished
 REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (25-FEB-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Mar 1, 1999 this sequence version replaced g1:3213168.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_DJ0530J23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information

about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send
 mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GSI-67A24, 200 bp overlap. Actual start of this clone is at base position 1 of RP4-530J23; actual end is at 7592 of GSI-67A24.

FEATURES

source	Location/Qualifiers
repeat_region	1..129837
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="7"
repeat_region	/map="7q21.1-q21.3"
repeat_region	/clone="RP4-530J23"
repeat_region	/clone_lib="RPI-4"
repeat_region	1..1172
repeat_region	/rpt_family="MALR"
repeat_region	1173..1541
repeat_region	/rpt_family="MALR"
repeat_region	1542..1962
repeat_region	/rpt_family="L1"
repeat_region	1980..2394
repeat_region	/rpt_family="L1"
repeat_region	2395..2582
repeat_region	/rpt_family="ALU"
repeat_region	2595..3425
repeat_region	/rpt_family="L1"
repeat_region	3426..3756
repeat_region	/rpt_family="ALU"
repeat_region	3757..3931
repeat_region	/rpt_family="L1"
repeat_region	3932..4095
repeat_region	/rpt_family="L1"
repeat_region	4365..4762
repeat_region	/rpt_family="Other"
repeat_region	5079..5262
repeat_region	/rpt_family="ALU"
repeat_region	5332..5358
repeat_region	/rpt_family="MER4-group"
repeat_region	5341..5804
repeat_region	/rpt_family="MER4-group"
repeat_region	6123..6217
repeat_region	/rpt_family="AT-rich"
repeat_region	6810..6911
repeat_region	/rpt_family="AT-rich"
repeat_region	6932..6960
repeat_region	/rpt_family="AT-rich"
repeat_region	6966..7050
repeat_region	/rpt_family="MER73-group"
repeat_region	7086..7423
repeat_region	/rpt_family="L2"
repeat_region	8602..8649
repeat_region	/rpt_family="L2"
repeat_region	8608..8687
repeat_region	/rpt_family="MIR"
repeat_region	8688..8924
repeat_region	/rpt_family="ALU"
repeat_region	9006..9146
repeat_region	/rpt_family="L2"
repeat_region	9147..9927
repeat_region	/rpt_family="MER73-group"
repeat_region	9928..10346


```

repeat_region      /rpl_family="L2"
                    11104..11413
                    /rpl_family="Alu"
                    complement(12342..12525)
                    /db.xref="GI:1113437"
STS
repeat_region      13093..13412
                    /rpl_family="L1"
                    14362..14549
                    /rpl_family="L1"
                    14553..14967
                    /rpl_family="L1"
repeat_region      15716..15747
                    /rpl_family="AT-rich"
repeat_region      15786..15821
                    /rpl_family="POLY-A"
repeat_region      17133..17361
                    /rpl_family="MIR"
                    17455..17841
                    /rpl_family="MALR"
repeat_region      17962..18047
                    /rpl_family="L2"
                    18232..18295
                    /rpl_family="L1"
repeat_region      18537..18650
                    /rpl_family="MIR"
repeat_region      18651..18943
                    /rpl_family="Alu"
                    18944..18957
                    /rpl_family="MIR"
                    18958..19010
                    /rpl_family="(TA)n"
repeat_region      19011..19068
                    /rpl_family="MIR"
                    20552..20660
                    /rpl_family="L2"
                    20891..21251
                    /rpl_family="MALR"
                    22154..22270
                    /rpl_family="L2"
                    22376..22531
                    /rpl_family="L2"
                    22573..22637
                    /rpl_family="AT-rich"
                    23527..23702
                    /rpl_family="MERL-type?"
                    24527..24650
                    /rpl_family="MER94"
                    24950..25256
                    /rpl_family="Alu"
                    25634..25677
                    /rpl_family="(CA)n"
                    25830..25999
                    /rpl_family="MIR"
                    26413..26449
                    /rpl_family="AT-rich"
                    26664..26697
                    /rpl_family="(TA)n"
                    27087..27107
                    /rpl_family="AT-rich"
                    29530..29577
                    /rpl_family="G-rich"
                    33727..33882
                    /rpl_family="MIR"

```

Query Match 0.7%: Score 20; DB 9; Length 129837;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 693 attcagcctgctgctga 712
 ||||||||||||||||
 Db 128831 ATTCAFGCCAGCTGCTGA 128850

```

RESULT 74
AC087702          AC087702 130981 bp DNA HTG 11-FEB-2001
LOCUS             Trypanosoma brucei clone RPI93-4512, *** SEQUENCING IN PROGRESS
DEFINITION        ***, 1 ordered pieces.
ACCESSION         AC087702
VERSION           AC087702.2 GI:12745096
KEYWORDS          HTG; HTGS; PHASE2.
SOURCE            Trypanosoma brucei.
ORGANISM          Trypanosoma brucei
                  Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE         1 (bases 1 to 130981)
AUTHORS           Lacombe D.J., El-Sayed, N.M., Kaul, S., Manless, D., Turner, C.M.R. and
                  Donelson, J.E.
TITLE             Analysis of a donor VSG gene and its expression site in African
                  trypanosomes
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 130981)
AUTHORS           El-Sayed, N.M. and Khalak, H.
TITLE             Direct Submission
COMMENT           Submitted (16-JUN-2001) The Institute for Genomic Research, 9712
                  Medical Center Dr., Rockville, MD 20850, USA
                  On Feb 11, 2001 this sequence version replaced gi:12232495.
                  * NOTE: This is a 'working draft' sequence. It currently
                  * consists of 1 contigs. Gaps between the contigs
                  * are represented as runs of N. The order of the pieces
                  * is believed to be correct as given. However the sizes
                  * of the gaps between them are based on estimates that have
                  * provided by the submittor.
                  * This sequence will be replaced
                  * by the finished sequence as soon as it is available and
                  * the accession number will be preserved.
                  1 130981: contig of 130981 bp in length.
FEATURES
source            1..130981
                  /organism="Trypanosoma brucei"
                  /isolate="GUT410.1"
                  /db_xref="taxon:5691"
                  /clone="RPI93-4512"
BASE COUNT        39063 a 28962 c 29199 g 33757 t
ORIGIN

```

Query Match 0.7%: Score 20; DB 2; Length 130981;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1118 aaacaacacaacataa 1137
 ||||||||||||||||
 Db 28755 AAAACACACACACATMAA 28774

```

RESULT 75
AC008370/c        AC008370 132171 bp DNA INV 03-AUG-1999
LOCUS             Drosophila melanogaster, chromosome 2R, region 44B-44C, BAC clones
DEFINITION        BACR09M11 and BACR44A15, complete sequence.
ACCESSION         AC008370 AC007302 AC007530
VERSION           AC008370.1 GI:5685882
KEYWORDS          HTG.
SOURCE            fruit fly.
ORGANISM          Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE         1 (bases 1 to 132171)
AUTHORS           Gelniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
                  Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
                  Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
                  Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
                  Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Ling, H.,
                  Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,

```

TITLE JOURNAL REFERENCE AUTHORS

Pfeiffer, B., Poon, L., Richards, S., Sequeira, A., Sethi, H.,
Svitskas, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M.,
Zieran, L.L. and Rubin, G.M.
Sequencing of Drosophila chromosome 2R, region 44B-44C
Unpublished
2 (bases 1 to 132171)

Celniker, S.E., Agbayani, A., Arcalata, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Ling, H.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Richards, S., Sequeira, A., Sethi, H.,
Svitskas, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M.,
Zieran, L.L. and Rubin, G.M.

TITLE
JOURNAL
COMMENT
Submitted (03-AUG-1999) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
On Aug 3, 1999 this sequence version replaced gl:5656695
gl:5656694.

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgpefruitfly@berkeley.edu.

FEATURES source

Location/Qualifiers
1..132171
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="44B-44C"
/clone="BACS BACR09N11 (D578) and BACR40A15 (D579)"
/clone_lib="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACe3.6)"
/note="BACR09N11 (D578) and BACR40A15 (D579) were
completed as a project. BACR09N11 extends from a minimal
overlap with its proximal neighbor DS02141 (D388) and
extends to a minimal overlap with its distal neighbor D363
at its BAC end at bp 132,171. Sequence from BACR40A15
was used to complete this region."

BASE COUNT 36580 a 28963 c 29822 g 36806 t
ORIGIN

Query Match 0.7%; Score 20; DB 3; Length 132171;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2499 tgcgaatgcagtgaaatcc 2518
|||||
DB 14519 TGCAATGCAGTGAATCC 14500

Search completed: March 25, 2002, 21:27:12
Job time: 34896 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 15:27:50 ; Search time 2764.97 Seconds
(without alignments)
11939.020 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcataagaacaa.....ctttaactgaactcgtc 3072

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_prod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	551	17.9	552	13	AO309404 CITHI-E1-
C 2	446	14.5	480	10	AV719179
C 3	342	11.1	364	10	A1263294
C 4	291	9.5	371	10	AV656315
C 5	223	7.3	261	10	AW337918
C 6	170	5.5	553	13	AO320928
C 7	168	5.5	476	11	BG210375
C 8	146	4.8	630	13	AO112439
C 9	136	4.4	509	10	AW418826
C 10	106	3.5	470	13	AO624020
C 11	98	3.2	412	11	H25984
C 12	93	3.0	499	10	A1023795

13	58	1.9	219	13	AO283386
14	55	1.8	635	11	B58691
15	55	1.8	840	11	BF207840
16	32	1.0	446	10	A1222422
17	30	1.0	251	11	BF903662
18	28	0.9	404	11	AO889169
19	23	0.7	606	11	BF829853
20	22	0.7	256	10	AV268403
21	22	0.7	437	13	A2740007
22	21	0.7	163	13	A2086396
23	21	0.7	343	10	AA784099
24	21	0.7	343	10	AA821339
25	21	0.7	371	10	A1645009
26	21	0.7	432	13	A2233093
27	21	0.7	437	13	AO555443
28	21	0.7	462	13	CNS00SV2
29	21	0.7	602	11	A2360053
30	21	0.7	1147	11	BG571228
31	20	0.7	64	11	B1142462
32	20	0.7	73	10	BE638399
33	20	0.7	205	13	AO069062
34	20	0.7	215	11	B131256
35	20	0.7	216	10	AA493806
36	20	0.7	225	10	A1366637
37	20	0.7	248	10	A1603832
38	20	0.7	299	13	A2891870
39	20	0.7	332	10	AW357412
40	20	0.7	336	13	AO952434
41	20	0.7	342	10	AA670222
42	20	0.7	380	13	A2991565
43	20	0.7	393	10	AA294641
44	20	0.7	393	10	AW838955
45	20	0.7	415	10	AA294540
46	20	0.7	415	11	T81898
47	20	0.7	439	10	AA294143
48	20	0.7	483	13	AO209355
49	20	0.7	491	11	BF601894
50	20	0.7	498	10	A1643174
51	20	0.7	505	11	BF602580
52	20	0.7	519	11	BG382487
53	20	0.7	521	10	A1670486
54	20	0.7	521	13	FR0031506
55	20	0.7	526	11	B1394563
56	20	0.7	535	11	BG351255
57	20	0.7	541	13	A2221490
58	20	0.7	554	13	A2296434
59	20	0.7	557	13	AO270247
60	20	0.7	571	11	BF258009
61	20	0.7	577	11	BG385206
62	20	0.7	577	13	AO570315
63	20	0.7	580	13	A2987277
64	20	0.7	605	11	BG098273
65	20	0.7	620	13	A2802117
66	20	0.7	627	10	A1511575
67	20	0.7	633	13	A2510047
68	20	0.7	644	10	AW874886
69	20	0.7	652	10	BE346038
70	20	0.7	653	10	A1317901
71	20	0.7	656	10	BE363495
72	20	0.7	660	11	BG310513
73	20	0.7	726	13	CNS02935
74	20	0.7	731	11	B65395
75	20	0.7	731	11	BG250425
76	20	0.7	734	13	AO329189
77	20	0.7	760	10	A1591096
78	20	0.7	772	11	BG310534
79	20	0.7	779	11	BG350988
80	20	0.7	783	11	B1184833
81	20	0.7	784	11	BF824698
82	20	0.7	891	11	BG248945
83	20	0.7	937	13	CNS02507
84	20	0.7	939	13	CNS02105
85	20	0.7	1036	11	BG176063

AO283386	RPCT11-78
B58691	CIT-HSP-201
BF207840	601862546
A1222422	qH04F06.x
BF903662	IL2-MT018
AO889169	HS-2161_B
BF829853	MR3-HN005
AV268403	AV268403
A2740007	RPCT-24-1
A2086396	RPCT-23-2
AA784099	d1908a1.f
AA821339	v6568a01.r
A1645009	v6568a01.y
A2233093	RPCT-23-9
AO555443	HS-5224_A
CNS00SV2	Arabidops
A2360053	1M0103H11
BG571228	602592168
B1142462	SMOV3MCA
BE638399	SMOVL2CAS
AO069062	HS-2255_B
B131256	dan96e06.
AA493806	nh02f04.s
A1366637	SMOV3MCA
A1603832	SMOV3MCA
A2891870	RPCT-24-1
AW357412	40862 MAR
AO952434	Sheared D
AA670222	ad19h11.s
A2991565	2M0275B24
AA294641	SMOV3MCA
AW838955	CM2-LT006
AA294540	SMOV3MCA7
T81898	Yd29a11.s1
AA294143	SMOV3MCA1
AO209355	HS-3240_A
BF601894	266590 MA
A1643174	v646909.y
BF602580	267845 MA
BG382487	296398 MA
A1670486	SMOV3MCA
FR0031506	Fugu rubr
B1394563	ppp1n.pk0
BG351255	088D06 MA
A2221490	GM-UMD001
A2296434	RPCT-23-1
AO270247	HS-2045_A
BF258009	HVSMF001
BG385206	306845 MA
AO570315	HS-5352_B
A2987277	2M0269115
BG098273	EST462792
A2802117	2M0061H04
A1511575	SMOVL3CAN
A2510047	1M0354H14
AW874886	SMOV3MCA
BE346038	SMOVL2CAS
A1317901	SMOV3MCA
BE363495	SMOVL2CAS
BG310513	SMOV3MCA
A1868746	Tetradon
B65395	CIT-HSP-202
BG250425	602362580
AO329189	nbx00044f
A1591096	tw91b09.x
BG310534	SMOV3MCA
BG350988	099C11 Ma
B1184833	UNC-P-FN-
BF824698	SMOV3MCA
BG248945	602361385
A1818456	Tetradon
A176270	Tetradon
BG176063	602335289

```
c 86 20 0.7 1396 11 BG254706
c 87 19 0.6 207 10 AV365606
c 88 19 0.6 209 10 AV144358
c 89 19 0.6 238 10 AV371617
c 90 19 0.6 256 10 AV1340263
c 91 19 0.6 272 11 BG381720
c 92 19 0.6 283 10 AV048726
c 93 19 0.6 304 13 AV094693
c 94 19 0.6 314 13 BS18669
c 95 19 0.6 332 10 BS18669
c 96 19 0.6 334 13 AV0285454
c 97 19 0.6 351 13 AV274848
c 98 19 0.6 355 10 AV356070
c 99 19 0.6 355 11 H88536
c 100 19 0.6 360 13 AV039398
```

ALIGNMENTS

```
RESULT 1
LOCUS AO309404/c 552 bp DNA GSS 22-DEC-1998
DEFINITION CITBI-E1.2528J13.TF CITBI-E1 Homo sapiens genomic clone 2528J13,
DNA sequence.
ACCESSION AO309404
VERSION AO309404.1 GI:4041438
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other GSS: CITBI-E1-2528J13.TR
CONTACT: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigf.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tadb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
```

FEATURES

```
source
1..352
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2528J13"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAcl; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
BASE COUNT 144 a 121 c 130 g 157 t
ORIGIN
```

```
Query Match 17.9%; Score 551; DB 13; Length 552;
Best Local Similarity 100.0%; Pred. No. 3.9e-267;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1427 aaatggttcattcggaacattacatcatcagcctgctcgtgacacgtg 1486
|||||
DB 552 AAATGGTTTCATTTCGACATACATCACTATACAGAGCTCCGTCGACACTGTG 493
```

```
OY 1487 gtcacatctggaagccaccagagcgtgtatgaagcacctcgcagcagtgatcaacacg 1546
|||||
DB 492 GGTATCTGTGGAAAGCCACCAGGCGCTTTATGAAGCACCTCGACGAGTATCAACACG 433
OY 1547 gctgccttcgcagcattccatcgccaaagagcctctctgtgagacaggaattctgcaa 1606
|||||
DB 432 GCTGCTTCTTCGAGCTTCCATCCCAAGAGCGCTCTCTGAGACAGAACTTTGGCAA 373
OY 1607 gtggaagaaacaccacgcagcaagaattctgaagcctaataatcattcctgtag 1666
|||||
DB 372 GTGGAAAACACCACGACGACGAATAATCTGAAGGCCATTAACATCAATTCCTTGTAG 313
OY 1667 agtctgcatccatttatcaagaagagatcatccaatcagccctagcagaagaattg 1726
|||||
DB 312 AGTGTGCATTCATTTATATCAAGAGATACATCAATCAATCAGCCCTGAGCAAGAAATTG 253
OY 1727 aagcttcttcacaggaataaagcttatatacaactcaggaacatcccgattctat 1786
|||||
DB 252 AAGCTTCTTTCAGGCTAAAGCTTATATCAACTCAGGGAACATCCCGATTACTTAT 193
OY 1787 ttgactctcttggaacatttgcacatttgcgaagtgcgaagtctcgaattcattcactgact 1846
|||||
DB 192 TTGACTTCTTTGAACATTTGCCCATTTGTCACAGTCTCTGACTTATTAACCTGACT 133
OY 1847 ttatggggagagctatgcttcatggaagaaagctgcagaagacacagtggaatccaca 1906
|||||
DB 132 TTTATGGGGAGGAGTATGCTTCATGGAAGAAAGCTGCAGAAAGACAGTGGAAATCCACA 73
OY 1907 tgaagagagcccaagaaactcaatcccaagcagagcgtatctgttcttcaactgga 1966
|||||
DB 72 TGGAAAGAGGCCCGCAAGAACTACATTCACAGAGGCTGTATCTTGTCTCAACTGGA 13
OY 1967 agcaggaattc 1977
|||||
DB 12 AGCAGGAATTC 2
```

```
RESULT 2
LOCUS AV719179/c 480 bp mRNA EST 16-OCT-2000
DEFINITION AV719179 GLC Homo sapiens cDNA clone GLCE0A10 5', mRNA sequence.
ACCESSION AV719179
VERSION AV719179.1 GI:10816331
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

REFERENCE

```
1 (bases 1 to 480)
Qian,B., Wu,T., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Y., Gu,Y., Chen,Z., and Han,Z.
Homo sapiens cDNA GLC clones
Unpublished (2000)
CONTACT: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
```

FEATURES

```
source
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCE0A10"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
```

```
/note="Vector: pbluescript sk(-): Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      138 a      120 c      85 g      137 t
ORIGIN
Query Match      14.5%; Score 446; DB 10; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.8e-214;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2605 catgaactgatacagagatacagctgctagaaacagctaccgactgactgctgctcctg 2664
      |||||||
DB 480 CATGAACCTGATCGACAGATGAACTGCTAGAAACACTCACCCTCATGATGCTGCCCTGG 421
QY 2665 gactgagatgcaagcagcctgagcaacccgtgttgaacatttggagaggtcccaaa 2724
      |||||||
DB 420 GCTCTGACGCTCCAGCGACCGCTGACACCCCTGTGAAACATTTGGAGAGAGGTCCACAA 361
QY 2725 ctcgtcaagcttggtgtgaanaactgagactacagatacagaagattagaatttagt 2784
      |||||||
DB 360 CTCGTCAACCTTGCTTGAAAACTGGACATCCACAGATACAGATTGAAATTTTAGGT 301
QY 2785 gcaatttttggaaaagaaacccctctgaaaactccagcaattgaattggcgggaaatcgt 2844
      |||||||
DB 300 GCATTTTGGAAAGAACCCCTCTGAAAACTTCACAGCACTTGAAATTTGGCGGGAATCGT 241
QY 2845 gtgagcagtgatgagatggtgcttcctcatggtgtatattgagaatccttaagcaattag 2904
      |||||||
DB 240 GTGACGAGATGATGATGCTTCCTTCATGAGGTGTATTTGAAACATCTTAAGCATTTAGTG 181
QY 2905 ttttttgaacttaagtaagaatcttcaactgatacagcaattagtcagaanaacttagc 2964
      |||||||
DB 180 TTTTGTGACTTATAGTACTAAGATTTCTACCTGATCCAGCATTAATCAGAAAACTTAGC 121
QY 2965 caagtgatatacgaatttaactttctcgaagaagctaggtctgttgggtgcaattgat 3024
      |||||||
DB 120 CAAGGTATTCGAAGTAACTTTCTGCAAGAGTAGGCTGTGGTGCAATTTGAT 61
QY 3025 gatgatgatactcagtgattatcagg 3050
      |||||||
DB 60 GATGATGATCTCAGTGTATTTATACAGG 35

RESULT 3
LOCUS      A1263294      364 bp      mRNA      EST      03-FEB-1999
DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
ACCESSION A1263294
VERSION A1263294.1 GI:3871497
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 364)
          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-ri@mail.nih.gov
          Life Technologies catalog #: 11548-013
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www.bio.llnl.gov/bbrp/image/image.html
          Insert length: 2146 Std Error: 0.00
          Seq primer: -40up from Gibco
          High quality sequence stop: 364.
          Location/Qualifiers
            1..364
            /organism="Homo sapiens"

FEATURES
source
```

```
/db_xref="taxon:9606"
/clone="IMAGE:2005417"
/clone_lib="NCI_CGAP_Pan1"
/clone_type="adenocarcinoma"
/tissue_type="DH10B"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: PCMV-SPOB6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT      117 a      84 c      55 g      108 t
ORIGIN
Query Match      11.1%; Score 342; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.9e-161;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2731 aagcttggttgaanaaacctgagagactcacagatacagagattagaattttagtgcat 2790
      |||||||
DB 364 AACCTTGGGTTGAAAACTGGAACCTCAGACATCAGATACAGATTGATTTAGTGCATTT 305
QY 2791 ttggaaaagaaacccctctgaaaactccagcaattgaattggcgggaaatcgttgagc 2850
      |||||||
DB 304 TTGGAAAGAACCCCTCTGAAAACTTCAGCACTTGAAATTTGGCGGGAATCGTGAGC 245
QY 2851 agtgatgagtgcttgccttcacatggtgtatttggaaatcctaagaattagtgctttt 2910
      |||||||
DB 244 AGTGATGATGATGCTTCCTTCATGAGGTGTATTTGAAACATCTTAAGCATTTAGTGT 185
QY 2911 gactttagtactaaagaatcttcaactgatacagcaattagtcagaanaacttagcaagt 2970
      |||||||
DB 184 GACTTATAGTACTAAGATTTCTACCTGATCCAGCATTTAGTGCAGAACTTAGCCAAAGTG 125
QY 2971 ttatccaagtaactttctcgaagaagctaggtctgttgggtgcaattgatgatgat 3030
      |||||||
DB 124 TTATCCAAGTAACTTTCTGCAAGAGTAGGCTGTGGTGCAATTTGATGATGAT 65
QY 3031 gatctcagtgattatcaggtgctttaaactagtagtgc 3072
      |||||||
DB 64 GATCTCAGTGTATTTACAGGTGCTTTTAAACTAGTATGCT 23

RESULT 4
LOCUS      AV656315      371 bp      mRNA      EST      07-SEP-2000
DEFINITION AV656315 GLC Homo sapiens cDNA clone GLCQD10 3', mRNA sequence.
ACCESSION AV656315
VERSION AV656315.1 GI:9877329
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 371)
          Qian,B., Wu,T., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
          Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
          ,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
          Homo sapiens cDNA clone
          Unpublished (2000)
          Contact: Zeguang Han
          Unpublished (2000)
          351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
          201203, P. R. China
          Tel: 86-21-50801919(ex.45)
          Fax: 86-21-50801922
          Email: hanzg@chgc.sh.cn
          This clone is available at CHGC in Shanghai.
          Location/Qualifiers
            1..371
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="GLCQD10"
```

```
/clone_lib="GIC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      112 a      85 c      91 g      82 t      1 others
ORIGIN

Query Match      9.5%; Score 291; DB 10; Length 371;
Best Local Similarity 99.7%; Pred. No. 1.2e-135;
Matches 341; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1873 gaaaagctgcagagaagacacagctgcaatccacatggaagagcccaaacctacatt 1932
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GAAAGGCTGCAGAAAGACACAGGTGGAATCCACATGGAAGAGCCCGAAGAACTACATT 60
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1933 cccagcagagctgtatcttctgcttcaactggaagcagaactcaggaactctggaagtc 1992
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 CCAGCAGGGGCTGTATCTTTGTTCTTCACTGGAAGCAGGAAATTCAGACTCTGAGAGTC 120
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1993 acactccgggatttcagcaagctgcaatgaagatacatatctcggggaataatttc 2052
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 ACACCTCCGGATTTCACCAAGTTCAATTAAGCAAGATATTCAGATATCTGGGAAATATTC 180
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2053 agctctgccacaagcctcagagctgcaataaagagatgctggtgctgagctggaagctc 2112
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AGCTCTGCCAAGCCTCAGGCTGCAATTAAGATGTGCTGTGTGGCTGGAAGCCTC 240
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2113 agtttgctctcagcagcttgaagaacattatctctcatggtggaagcagtcacctc 2172
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 AGTTTGCTCTCAGCAGCTGTAAAGACATTTATCTCTCAAGTGGAAGCCAGTCCTC 300
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2173 accataaagatgagagagcaccatcacatctgtaacaacctg 2214
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 ACCATTAAGATGAGAGGCACATCATCTGTATACAAACCTG 342
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5
AM337918/c 261 bp mRNA EST 31-JAN-2000
LOCUS he12h11.x1 NCI_CGAP_CML1 Homo sapiens cDNA IMAGE:2918853 3'
DEFINITION mRNA sequence.
ACCESSION AM337918
VERSION AM337918.1 GI:6834544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 261)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -400p from Gibco
High quality sequence stop: 201.
Location/Qualifiers
1. .261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2918853"
```

```
/clone_lib="NCI_CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-Sport6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."

BASE COUNT      86 a      54 c      38 g      83 t
ORIGIN

Query Match      7.3%; Score 223; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.8e-101;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 cagtgatgagtggttccttcacatgagtgatggaatcagaactagatgtttt 2909
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 261 CAGTGATGAGTGCTTCCTTCATGAGGTGATTTGAACAATCTTAAGCAATAGGTTT 202
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2910 tgaattagtaactaagaattctcactgataccagatgacagaaactagaact 2969
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 201 TGACTTTAGTACTAAGAATTTTCTACTGATCCAGATTTAGTGAAGAACTTAGCCAGT 142
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2970 gttatccaagttaactcttcctgcagaagcagctggtggtggaattgatgta 3029
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141 GTTATCCAGATTAACTTTCTGCAAGAGCTAGGCTGTGGGTGGCAATTTGATGATGA 82
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 3030 tgatctcagtgatctacagtgcttctaactagtaactgct 3072
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 81 TGATCTCAGTGTTTATTCAGGTCCTTTTAACATGATGATGCT 39
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 6
AQ320928/c 553 bp DNA GSS 06-MAY-1999
LOCUS RPC111-93C9.TV RPC1-11 Homo sapiens genomic clone RPC1-11-93C9, DNA
DEFINITION sequence.
ACCESSION AQ320928
VERSION AQ320928
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 553)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Sub,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPC111-93C9.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Library availability, please contact Pieter de Jong
(pietere@jeng.med.buflalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buflalo.edu/ordering) or from
Research Genetics (http://info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. .553
/organism="Homo sapiens"
/db_xref="GDB:7535384"
/db_xref="taxon:9606"
/clone_lib="RPC1-11-93C9"
/clone_lib="RPC1-11"
```

/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site.1: ECORI; Site.2: ECORI;
RFLII Human Male BAC Library"
BASE COUNT 170 a 107 c 114 g 162 t
ORIGIN

Query Match 5.5%; Score 170; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.8e-74;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2615 tcgaacgagtagacgtgctagacagctcaccgcactgctgcgcctgggctgtgacg 2674
|||||
DB 258 tcgacagagtagacgtgctagacagctcaccgcactgctgcgcctgggctgtgacg 199
|||||
QY 2675 tcgaacgagctgagcagcagctgttgaacattggaggaggtcccaactgcgtcaagc 2734
|||||
DB 198 tcgaagggcagcctgagcagcagctgttgaacattggaggaggtcccaactgcgtcaagc 139
|||||
QY 2735 ttgggttgaacaaactgagacacagatcacagagattggaatttagt 2784
|||||
DB 138 ttgggttgaacaaactgagacacagatcacagagattggaatttagt 89
|||||

RESULT 7
LOCUS BG210375 476 bp mRNA EST 21-APR-2001
DEFINITION R5729913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 476)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velooso,N., Hess,J., Cochren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 360.
Location/Qualifiers
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

TITLE
JOURNAL
COMMENT
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 360.
Location/Qualifiers
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

FEATURES
source
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

BASE COUNT 120 a 99 c 107 g 149 t
ORIGIN
Query Match 5.5%; Score 168; DB 11; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.8e-73;
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2873 tgggtatattgagatcttaagcaatagtgctttttagacttagtactaaagaatttc 2932

|||||
DB 242 tgggtatattgagatcttaagcaatagtgctttttagacttagtactaaagaatttc 301
|||||
QY 2933 taactgattccagcaatagtcagaacaaactagccaaagtgttaccagtaactttctgc 2992
|||||
DB 302 taccgtatccagcaatagtcagaacaaactagccaaagtgttaccagtaactttctgc 361
|||||
QY 2993 aagaagctagctgttgggtgagcaatttgatgatgatctcagtg 3040
|||||
DB 362 aagaagctagctgttgggtgagcaatttgatgatgatctcagtg 409
|||||

RESULT 8
LOCUS AQ112439 630 bp DNA GSS 29-AUG-1998
DEFINITION CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA
sequence.
ACCESSION AQ112439
VERSION AQ112439.1 GI:3484599
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 630)
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other GSSs: CIT-HSP-2372C1.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamad@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"

TITLE
JOURNAL
COMMENT
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamad@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"

FEATURES
source
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 188 a 126 c 124 g 192 t
ORIGIN
Query Match 4.8%; Score 146; DB 13; Length 630;
Best Local Similarity 100.0%; Pred. No. 2.5e-62;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2349 agctaaagcctcgaaacacctgaagaagatggttattcatcttgaccacattgtctga 2408
|||||
DB 485 agctaaagcctcgaaacacctgaagaagatggttattcatcttgaccacattgtctga 544
|||||
QY 2409 catltagagaggaatcgatcatatagtcgaagctctgtcaagtgaaacctgagacctga 2468
|||||
DB 545 catltagagaggaatcgatcatatagtcgaagctctgtcaagtgaaacctgagacctga 604
|||||
QY 2469 agaattcaattagctcctcctgctct 2494
|||||
DB 605 agaattcaattagctcctcctgctct 630
|||||

RESULT 9
LOCUS AM418826/c
DEFINITION ha21e11.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2874380 3',
mRNA sequence.
ACCESSION AM418826
VERSION AM418826.1 GI:6946758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 509)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/dbpr/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 461.
Location/Qualifiers
source 1..509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonoids 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Falcima Bernaldo."

BASE COUNT 161 a 99 c 80 g 169 t
ORIGIN

Query Match 4.4%; Score 136; DB 10; Length 509;
Best Local Similarity 100.0%; Pred. No. 2.8e-57;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2937 tgaccagcattagtcagaaacttagcgaagtgtatccaagtaactttctgcaga 2996
|||||
Db 509 TGATCCAGCATTTACTGAGAAACTTAGCCAAAGTGTATCCAAAGTATTTCGCAAGA 450
|||||

QY 2997 agctaggcttgcttggtgcgaatttgatgacatcagtgctatcagaagtgtctt 3056
|||||
Db 449 AGCTAGGCTGTGTGGTGGCAATTTGATGATGATGATCTGCTGTTTACAGCTGCTTT 390
|||||

QY 3057 taaactagtaactgct 3072
|||||
Db 389 TAAACTAGTAAGTCTGCT 374
|||||

RESULT 10
LOCUS AO624020 470 bp DNA GSS 16-JUN-1999

DEFINITION HS_5378_B2.C12.SPEE RPCT-11 Human Male BAC library Homo sapiens
genomic clone Plate=954 Col=24 Row=F, DNA sequence.
ACCESSION AO624020
VERSION AO624020.1 GI:5086412
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 470)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,T., Young,T., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC
library availability, please contact Pieter de Jong
(piet@redj.org.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (inforesgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 954 Row: F Column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 470.
Location/Qualifiers
source 1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=954 Col=24 Row=F"
/clone_lib="RPCT-11 Human Male BAC library"
/sex="male"
/note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 142 a 101 c 107 g 117 t 3 others
ORIGIN

Query Match 3.5%; Score 106; DB 13; Length 470;
Best Local Similarity 100.0%; Pred. No. 4.2e-42;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2034 atatctggggaataatctcagctctgccacaagctcagctgcaataaagagatgtgc 2093
|||||
Db 64 ATATCTGGGGAATAATATTCAGCTCTGCCACAAGCTCAGCTGCAATAAAGAGATGTGC 123
|||||

QY 2094 tgggtgctgcgggaagcctcagtttgctccagacactgtaagaac 2139
|||||
Db 124 TGGTGTGGCTGGAGAGCCTCAGTTGGTCTCAGCAGCCTGTAGAGAC 169
|||||

RESULT 11
LOCUS H25984/c 412 bp mRNA EST 10-JUL-1995
DEFINITION Y136g07.r1 Soares breast 3NDBst Homo sapiens cDNA clone
IMAGE:162300 5', mRNA sequence.
ACCESSION H25984
VERSION H25984.1 GI:895107
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 412)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 942
High quality sequence stops: 347
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 942 Std Error: 0.00
Seq primer: M13Rpi
High quality sequence stop: 347.
Location/Qualifiers
1. 412
/organism="Homo sapiens"
/db_xref="GDB:576544"
/db_xref="taxon:9606"
/clone="IMAGE:162300"
/clone_lib="Soares breast 3MBHst"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAGAGTGGACGCGCCCTTTTGTGATGATGATGATC 281 (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot - 20. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
129 a 84 c 73 g 124 t 2 others

Query Match
Best Local Similarity 100.0%; Pred. No. 4.6e-38;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
LOCUS
DEFINITION
ACCESSION
KEYWORDS
SOURCE
ORGANISM

AI023795 499 bp mRNA EST 28-AUG-1998
ox08d03.x1 Soares fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
clone IMAGE:1655717 3', mRNA sequence.
AI023795 GI:3238839
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 499)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1179 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 439.
Location/Qualifiers
1. 499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1655717"
/clone_lib="Soares-fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subcloned version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGACAGATTAATTAAGATCTTTTGTGATGATGATGATC 322 (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
144 a 109 c 98 g 148 t

Query Match
Best Local Similarity 100.0%; Pred. No. 1.6e-35;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AO283886 219 bp DNA GSS 27-APR-1999
RPC111-78E13.TV RPC1-11 Homo sapiens genomic clone RPC1-11-78E13,
DNA sequence.
AO283886
AO283886.1 GI:3910204
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 219)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P., and Venter, J.C.
Use of Human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC

A1222422 446 bp mRNA EST 30-NOV-1998
 LOCUS qh04f06.x1 Soares_NFL_T.GBC.S1 Homo sapiens CDNA clone
 DEFINITION IMAGE:1843715 3', mRNA sequence.
 ACCESSION A1222422
 VERSION A1222422.1 GI:3804625
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 446)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-rt@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 546 Std Error: 0.00
 Seq primer: -40up from Gibco
 High quality sequence stop: 437.
 Location/Qualifiers
 1..446
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image:1843715
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site.1: Not I; Site.2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBH19W, testis NHT, and B-cell
 NCI-CGAP-GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 1 M.A.G.E. clones 297480-302087, 662632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo.
 Soares 95 c 86 g 137 t
 BASE COUNT 128 a
 ORIGIN
 Query Match 1.0%; Score 32; DB 10; Length 446;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 263 gctcttcacacagacagaagagacttg 294
 ||||||||||||||||||||||||||||||||
 Db 415 gtccttttcacacagacagaagagacttg 446
 RESULT 17
 LOCUS BF903662 251 bp mRNA EST 18-JAN-2001
 DEFINITION IL2-MT0180-181200-276-F03 MT0180 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BF903662
 VERSION BF903662.1 GI:12295121
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 251)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zagó, M.A., Bordão, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?IL2-IL2&IL2-MT0180-181200-276-F03&IL2-2000-12-18&IL2-1>)
 Seq primer: puc 18 forward
 High quality sequence start: 45
 High quality sequence stop: 96.
 Location/Qualifiers
 1..251
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MT0180"
 /dev_stage="Adult"
 /note="Organ: marrow; Vector: puc18; Site.1: SmaI; Site.2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 72 a 62 c 56 g 61 t
 ORIGIN
 Query Match 1.0%; Score 30; DB 11; Length 251;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1057 gagttccactctcacacacacacacgctg 1086
 ||||||||||||||||||||||||||||||||
 Db 52 GAGTCCACTCTCACACACACACACGCTG 81
 RESULT 18
 LOCUS AO889169 404 bp DNA GSS 10-NOV-1999
 DEFINITION HS_2161_B1_A01_TTC CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.
 ACCESSION AO889169
 VERSION AO889169.1 GI:6345359
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 404)
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 2161 row: B column: 1
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 404.

FEATURES

Source

1. 404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="plate=2161 Col-1 Row-B"
 /note="Organ: sperm; Vector: pBelOBAC1; BAC clones in E-Coli DH10B"
 /sex="male"
 BASE COUNT 119 a 93 c 70 g 122 t
 ORIGIN

Query Match

Best Local Similarity 100.0%; Score 28; DB 13; Length 404;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 tagagggatcctcactgatttgaaa 180
 ||||||||||||||||||||||||||||
 Db 200 TAGAGGGATCATTCACATGATTGAAA 173

RESULT 19

BF829853 606 bp mRNA EST 13-JAN-2001
 LOCUS BF829853
 DEFINITION M3-HN0052-261200-002-c12 HN0052 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF829853
 VERSION BF829853.1 GI:12175855

SOURCE

human.
 ORGANISM

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 606)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M.Jr., Zago,M.A., Bordin,S., Costa,F.F., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE

JOURNALT. 2020263
 MEDLINE Contact: Simpson A.J.G.
 COMMENT Laboratory of Cancer Genetics
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

COMMENT

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=MR3&t2=MR3-HN0052-261200-002-c12&t3=2000-12-26&t4=1>
 Seq primer: puc 18 forward
 High quality sequence start: 24
 High quality sequence stop: 538.

FEATURES

Source

1. 606
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HN0052"
 /dev_stage="Adult"
 /note="Organ: head_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application

BASE COUNT

119 a 177 c 195 g 113 t 2 others
 ORIGIN

Query Match

Best Local Similarity 100.0%; Score 23; DB 11; Length 606;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1887 agacacagtgatgcacatg 1909
 ||||||||||||||||||||
 Db 100 AGACACAGTGATGCACATG 122

RESULT 20

AV268403 256 bp mRNA EST 05-NOV-1999
 LOCUS AV268403
 DEFINITION AV268403 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4930534P12 3', mRNA sequence.
 ACCESSION AV268403
 VERSION AV268403.1 GI:6256440

SOURCE

house mouse.
 ORGANISM

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 256)
 Konno,H., Alzawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata ,Y., Shigenoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tomioka,N., Tsunoda,Y., Watanabe,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Konno,H., et al. 1999)
 Unpublished (1999)

COMMENT

JOURNALT. 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/,
 Sasaki,N., Izawa,M., Watanabe,M., Okazaki,Y. and Hayashizaki ,Y.
 Mutsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki ,Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh,M., Kitsuana,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (<http://genome.ritc.riken.go.jp>) for further details.

FEATURES

Source

1. 256
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="4930534P12"
 /note="RIKEN full-length enriched, adult male testis

(DH10B)"
 /sex="male"
 /tissue.type="testis"
 /dev.stage="adult"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCTCGAGTAAATTAATATCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTAAATTAATTAATATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+)-after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3' end: BamHI."

BASE COUNT
 80 a 36 c 31 g 109 t

ORIGIN

Query Match 0.7%; Score 22; DB 10; Length 256;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2555 tctctgattatcagaanaatla 2576
 ||||||||||||||||||
 Db 45 TTCTGTGATTATCAGAAATTA 66

RESULT 21
 A2740007 437 bp DNA GSS 25-JAN-2001
 LOCUS RPCI-24-103A13.TV RPCI-24 Mus musculus genomic clone RPCI-24-103A13
 DEFINITION , DNA sequence.
 ACCESSION A2740007
 VERSION A2740007.1 GI:12510653
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 437)
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Aklnet,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-103A13.TJB
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 103 row: A column: 13
 Seq primer: 17
 Class: BAC ends.

FEATURES
 source
 1..437
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"

/clone="RPCI-24-103A13"
 /clone.lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT
 194 a 81 c 63 g 99 t

ORIGIN

Query Match 0.7%; Score 22; DB 13; Length 437;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1116 gaaacaacaacacacataaa 1137
 ||||||||||||||||||
 Db 204 GAAACAACAACACACATATAA 225

RESULT 22
 A2086396 163 bp DNA GSS 08-MAY-2000
 LOCUS RPCI-23-26D1.TV RPCI-23 Mus musculus genomic clone RPCI-23-26D1,
 DEFINITION DNA sequence.
 ACCESSION A2086396
 VERSION A2086396.1 GI:7728130
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 163)
 Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnet,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-26D1.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pdejong@med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 26 row: D column: 1
 Seq primer: 17
 Class: BAC ends.

FEATURES
 source
 1..163
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-26D1"
 /clone.lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J vector: kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ACCESSION	AA821339
VERSION	AA821339.1
KEYWORDS	GI:2891207
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 343)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellensberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R. The WashU-HMNI Mouse EST Project Unpublished (1996)
TITLE	Contact: Marra M/Mouse Est Project
JOURNAL	WashU-HMNI Mouse EST Project
COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNC, contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:624608 Seq primer: -28ml3 rev1 ET from Amersham High quality sequence stop: 336. Location/Qualifiers 1..343 /organism="Mus musculus" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:1151400" /clone_id="Stratagene mouse skin (#937313)" /sex="females" /tissue_type="whole skin" /dex_stage="11 weeks old" /lab_host="SOLR (kanamycin resistant)" /note="Organ: skin; Vector: pluscript SK.; Site:1: EcORI ; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAAATCGGACGAG 3' -3' adaptor sequence: 5' CTCGACGTTTTTTTTTTTTTTT 3'"
BASE COUNT	84 a 103 c 95 g 61 t
ORIGIN	
Query Match	0.7%; Score 21; DB 10; Length 343;
Best Local Similarity	100.0%; Pred. No. 40;
Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	752 atgaattcaagcccccagact 772
DB	319 ATGAAATCAAGCCCGACAGACT 339
RESULT	25
AI645009	
LOCUS	AI645009 371 bp mRNA EST 29-APR-1999
DEFINITION	v668a01.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
ACCESSION	AI645009
VERSION	AI645009.1
KEYWORDS	GI:4723484
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 371)
AUTHORS	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theisinger,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter

TITLE
JOURNAL
COMMENT

,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:624608

This read is a RESSEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 355.
Location/Qualifiers

FEATURES
source

```
1..371
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1151400"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI
Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo
dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb. Uni-ZAP XR Vector. ~5'
adaptor sequence: 5' GAAATTCGCGACGAG 3'-3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTT 3'"
```

BASE COUNT
ORIGIN

```
92 a 109 c 101 g 69 t
```

Query Match 0.7%; Score 21; DB 10; Length 371;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 752 atgaattcaagcccgagaact 772
|||||

Db 321 ATGAATTCAGCCCGAGAACT 341

RESULT 26
A2223093 432 bp DNA GSS 14-JUN-2000
LOCUS RPCI-23-94H20.TV RPCI-23 Mus musculus genomic clone RPCI-23-94H20,
DEFINITION DNA sequence.
ACCESSION A2223093
VERSION A2223093.1 GI:8531142
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 432)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinet
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-94H20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

TITLE
JOURNAL
COMMENT

FEATURES
source

Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 94 row: H column: 20
Seq primer: 77
Class: BAC ends.
Location/Qualifiers

```
1..432
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-94H20"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1:  
EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."
```

BASE COUNT
ORIGIN

```
75 a 123 c 122 g 112 t
```

Query Match 0.7%; Score 21; DB 13; Length 432;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 752 atgaattcaagcccgagaact 772
|||||

Db 115 ATGAATTCAGCCCGAGAACT 95

RESULT 27
A0555443 437 bp DNA GSS 29-MAY-1999
LOCUS HS.5224.A2.B05.SP6E RPCI-11 Human Male BAC library Homo sapiens
DEFINITION genomic clone Plate=800 Col=10 Row=C, DNA sequence.
ACCESSION A0555443
VERSION A0555443.1 GI:4914996
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 437)
Mahairas,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 800 row: C column: 10

Seq primer: SP6
Class: BAC ends
High quality sequence stop: 437.
Location/Qualifiers

FEATURES

1. 437

Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=800 Col=10 Row=C"
/clone.lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

BASE COUNT 168 a 60 c 81 g 128 t
ORIGIN

Query Match 0.7%; Score 21; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1670 gtggcattcattatcaag 1690
|||||
DB 387 GTGGCATTCTATATCAAG 367

RESULT 28
CNS005V2/c 462 bp DNA GSS 28-JUN-1999
LOCUS Arabidopsis thaliana genome survey sequence SP6 end of BAC T2N9 of
DEFINITION TMU library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.
ACCESSION AT088892
VERSION AT088892.1 GI:5290032
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 462)
AUTHORS Salanoubat,M., Choinsne,N., Artiguenave,F., Brothier,P., Wincker,P.,
Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 462)
AUTHORS
TITLE Genoscope.
JOURNAL Direct Submission
Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

FEATURES

Source

1. 462
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone.lib="TMU"
/clone="T2N9"
/note="end : SP6"
BASE COUNT 129 a 87 c 70 g 176 t
ORIGIN

Query Match 0.7%; Score 21; DB 13; Length 462;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 591 caaatgctctctcccg 611
|||||
DB 239 CAAATGCTCTCTCCG 219

RESULT 29
A2360053/c 602 bp DNA GSS 02-OCT-2000
LOCUS 1M0103H11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0103H11 F, DNA sequence.
ACCESSION A2360053
VERSION A2360053.1 GI:10473753
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 602)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10Kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: CGTGTGAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 602.
Location/Qualifiers

FEATURES

Source

1. 602
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0103H11"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab.host="E. coli strain XL10-Gold, 71-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1147321149b/AF12972.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 139 a 153 c 139 g 171 t
ORIGIN

Query Match 0.7%; Score 21; DB 13; Length 602;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1710 cctgagccaagaattgaac 1730
|||||
DB 21 CCTGAGCCCAAGATTGAAC 1

Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu

Seq primer: pbluescript SK.

FEATURES

Source

Location/Qualifiers

1..73

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone="SMOVL2CASI5B03"

/clone_1lb="Onchocerca volvulus L2 larvae cDNA

(SAM98MIM-OvL2)"

/dev_stage="L2"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. mRNA was

prepared from approximately 9,000 L2s isolated from

infected mosquitoes from Kumba, Cameroon and converted to

double-stranded cDNA using reverse transcriptase and

oligo(dT) followed by RNase H and DNA pol I. The library

has 7.3 x 10E4 independent recombinants and the average

insert size is approximately 1kb. The library was

constructed by Michelle Lizotte-Waniewski. The library is

available from Dr.S.A.Williams, email: genome@smith.edu."

22 a 11 c 19 g 21 t

BASE COUNT

22 a 11 c 19 g 21 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Length 73;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2329 gaagaagatgctataaact 2348

Db 27 GAAGACATGCTATATAAACT 46

RESULT 33
LOCUS A0069062/c

DEFINITION HS.2255.B2.H03.MR.CIT Approved Human Genomic Sperm Library D Homo

ACCSSION A0069062 GSS 04-AUG-1998

VERSION A0069062.1 GI:3384261

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 205)

MAHAIRES,G.G., WALLACE,J.C., SMITH,K., SWARTZELL,S., HOLZMAN,T.,

KEILER,A., SHAKER,R., FURLONG,J., YOUNG,J., ZHAO,S., ADAMS,M.D. and

HOOD,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

COMMENT Contact: Mahaires GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2255 row: P column: 6

Class: BAC ends

High quality sequence stop: 205.

FEATURES

Source

Location/Qualifiers

1..205

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate=2255 Col=6 Row=P"

/clone_1lb="CIT Approved Human Genomic Sperm Library D"

BASE COUNT

54 a 41 c 52 g 57 t 1 others

ORIGIN

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

Query Match

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Length 205;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 taagaagaaacacacacaa 1131

Db 134 TACGAAACAAACACACAAA 115

RESULT 34
LOCUS BI315256

DEFINITION BI315256 215 bp mRNA EST 20-JUL-2001

ACCSSION BI315256 dah9e06.y1 NICHD XGC Emb4 Xenopus laevis cDNA clone IMAGE:4957906

VERSION BI315256.1 GI:14989583

KEYWORDS EST.

SOURCE EST.

ORGANISM African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 215)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps@emall.nih.gov

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: infoimage.llnl.gov

Seq primer: -40R from Gibco

High quality sequence stop: 197.

FEATURES

Source

Location/Qualifiers

1..215

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:4957906"

/clone_1lb="NICHD XGC Emb4"

/dev_stage="embryo, stage 31-32"

/lab_host="DH10B (phage-resistant)"

/note="Organ: whole embryo; Vector: pCMV-SPORT6; Site_1:

NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo

dT. Average insert size 2.1 kb. Constructed by Life

Technologies. Note: This is a Xenopus Gene Collection (XGC

) library."

BASE COUNT 72 a 32 c 37 g 74 t

ORIGIN

Query Match 0.7%; Score 20; DB 11; Length 215;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Length 205;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2790 ttgtgaaagacccctga 2809

Db 196 TTTTGAAGAACCCCTGTA 215

RESULT 35

LOCUS AA493806 216 bp mRNA EST 19-AUG-1997

```

DEFINITION nh02f04.s1 NCI-CGAP_Thy1 Homo sapiens cDNA clone IMAGE:943135, mRNA
sequence.
ACCESSION AA493806
VERSION AA493806.1 GI:2223647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 216)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LML at:
www.bio.lml.gov/dbrip/image/image.html
Insert length: 294 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
Source
1..216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:943135"
/clone_lib="NCI CGAP_Thy1"
/tissue_type="thyroid"
/lab_host="DH10B"
/Note="Vector: pAMP10; mRNA made from invasive thyroid
tumor, cDNA made by oligo-dT priming. Non-directionally
cloned. Size-selected on agarose gel, average insert size
600 bp. Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT 57 a 65 c 37 g 57 t
ORIGIN
Query Match 0.7%: Score 20: DB 10: Length 216;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 2190 gccatcacatctgtaaca 2209
|||||
Db 92 gccatcacatctgtaacaa 111

RESULT 36
A1366637 225 bp mRNA EST 08-JAN-1999
LOCUS SMOV3MCM11C04SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SI96M1M-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCM11C04 5',
mRNA sequence.
ACCESSION A1366637
VERSION A1366637.1 GI:4136382
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 225)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith

```

```

College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
Source
1..225
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCM11C04"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SI96M1M-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRP"
/Note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 106 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigman@bc.org)."

BASE COUNT 65 a 39 c 57 g 64 t
ORIGIN
Query Match 0.7%: Score 20: DB 10: Length 225;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 2329 gaagaagatgctataaact 2348
|||||
Db 30 GAAGAAGATGCTATAAACT 49

RESULT 37
A1603832 248 bp mRNA EST 21-APR-1999
LOCUS SMOV3MCM27A08SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SI96M1M-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCM27A08 5',
mRNA sequence.
ACCESSION A1603832
VERSION A1603832.1 GI:4612981
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 248)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
Source
1..248
Location/Qualifiers

```

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM27A08"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA
(SI96M1W-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRP"
/Note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20x FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by Phase H
and DNA pol I. The library was constructed in the lambda
Uni-zap XR vector and has 1 x 106 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigmen@dc.org)."
BASE COUNT      71 a      47 c      58 g      69 t      3 others
ORIGIN

Query Match      0.7%; Score 20; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2329 gaagaagatgctataaact 2348
         |||||
Db      14 GAAGACATGCTATTAACACT 33

RESULT 38
AZ891870      299 bp      DNA      GSS      05-MAR-2001
LOCUS      RRCI-24-180D14.TJ RPCI-24 Mus musculus genomic clone RRCI-24-180D14
DEFINITION      , DNA sequence.
ACCESSION      A2891870
VERSION      A2891870.1 GI:13210815
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 299)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinet, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Georegeorgis, E.,
Russell, D., de Jong, P. and Fraser, C. M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-180D14.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@tigr.org). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
plate: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Size: 180 row: D column: 14
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers

```

```

source
1. 299
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-180D14"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/Note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      103 a      57 c      85 g      54 t
ORIGIN

Query Match      0.7%; Score 20; DB 13; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      954 tgaagcttgctgcacaa 973
         |||||
Db      202 TGAAGCTTGTGCTCCAAA 221

RESULT 39
AM357412      332 bp      mRNA      EST      25-APR-2001
LOCUS      AM357412
DEFINITION      40862 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      AM357412
VERSION      AM357412.1 GI:6861490
KEYWORDS      EST.
SOURCE      EST.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 332)
Smith, T. P. L., Grose, W. M., Freking, B. A., Roberts, A. J., Stone, R. T.,
Casas, E., May, J. E., White, J., Cho, J., Fahrenkrug, S. C., Bennett
G. L., Heaton, M. P., Laegreid, W. M., Rohrer, G. A., Chitko-McKown, C. G.,
Pette, G., Holt, I., Karayancheva, S., Liang, F., Quackenbush, J. and
Keeler, J. W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@e-mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 20
and -mismatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCACAGTCAGCAGC
Plate: 20 row: J column: 17
Seq primer: ATTGAGTGACACTATAG.
FEATURES
source
1. 332
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1lb="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: PCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

```

BASE COUNT 59 a 105 c 88 g 80 t
 ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 698 tggcagtcgtcgaactg 717
 ||||||||||||||||
 Db 77 TGCCATGCTCTGAACTG 96

RESULT 40
 A0952434/c 336 bp DNA GSS 27-JAN-2000
 LOCUS
 DEFINITION Sheared DNA-48N4.TR Sheared DNA Trypanosoma brucei genomic clone
 ACCESSION A0952434
 VERSION A0952434
 KEYWORDS
 SOURCE GSS.
 ORGANISM Trypanosoma brucei.
 Trypanosoma brucei.
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE
 1 (bases 1 to 336)
 El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
 Gerard, C., Leech, V., de Jong, P., Ullrich, E., Melville, S., Donelson, J.,
 Fraser, C. and Adams, M.
 Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 Unpublished (1999)
 Other GSS: Sheared DNA-48N4.TF
 CONTACT: Majib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
<http://www.tigr.org/cdb/mdb/tbdb/>.
 Seq primer: M13-Reverse
 Class: shotgun.

FEATURES
 source

1..336 Location/Qualifiers
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone_1lb="Sheared DNA-48N4"
 /clone_1lb="Sheared DNA"
 /note="Vector: pUC18; Site_1: SmaI. Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaubin and B. Bartell, Oxford University
 Press, 1999)."
 BASE COUNT 72 a 62 c 68 g 134 t
 ORIGIN

Query Match 0.7%; Score 20; DB 13; Length 336;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1118 aaacaacaacaataaa 1137

Db 171 AAAACAACACAAACATTA 152
 ||||||||||||||||

RESULT 41
 AA670222/c 342 bp mRNA EST 20-NOV-1997
 LOCUS
 DEFINITION ad19h11.s1 Soares-NbHFB Homo sapiens CDNA clone IMAGE:878757 3',
 mRNA sequence.
 ACCESSION AA670222
 VERSION AA670222
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 342)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 CONTACT: Wilson R
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through INM; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amerisham
 High quality sequence stop: 334.

FEATURES
 source

1..342 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="IMAGE:878757"
 /clone_1lb="Soares-NbHFB"
 /dev_stage="15 wk post natal"
 /lab_host="DH10B"
 /note="Organ: whole brain; Vector: p77/3D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) 3',
 15' AACTGGAGAAATTCGCGCCGCAATATTTTATTTTATTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p77/3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 72 a 122 c 81 g 67 t
 ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 613 ctcaagcagggcccaaggtg 632
 ||||||||||||||||
 Db 137 CTCACGAGGCCACGAGGTG 118

RESULT 42
 A2991565 380 bp DNA GSS 27-APR-2001
 LOCUS
 DEFINITION 2M0275B24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0275B24 R, DNA sequence.
 ACCESSION A2991565
 VERSION A2991565.1 GI:13862792
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 380)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 TITLE Mouse whole genome scaffolding with paired end reads from 10Kb
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0275 Row: B Column: 24
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 380.

FEATURES
 source
 1..380
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0275R24"
 /clone_lib="Mouse 10kb plasmid UUGC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42ny. Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g11473211419b/AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 140 a 71 c 97 g 72 t
 ORIGIN

Query Match 0.7%; Score 20; DB 13; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1118 aaaaacacacacacacataaa 1137
 ||||||||||||||||||||
 Db 35 AAAACAAACACAAACATMAA 54

RESULT 43
 AA294641 393 bp mRNA EST 12-NOV-1997
 LOCUS AA294641
 DEFINITION SMOV3MCA885SK Onchocerca volvulus molting L3 larva cDNA (SL96MLW-OVml3) Onchocerca volvulus cDNA clone Sml3CO885 5', mRNA sequence.
 ACCESSION AA294641
 VERSION AA294641.1 GI:2099635
 KEYWORDS -EST.

SOURCE
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 393)
 AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: phluescript SK.

FEATURES
 source
 1..393
 Location/Qualifiers
 1..393
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="Sml3CO885"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLW-OVml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda uni-zap XR; Site-1: Eco RI; Site-2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+NCYC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda uni-zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nyc.oxg.org)."

BASE COUNT 104 a 79 c 96 g 111 t 3 others
 ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 393;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2329 gaagaagatgctataaact 2348
 ||||||||||||||||||||
 Db 42 GAAGAGATGCTATTAACACT 61

RESULT 44
 AW838955 393 bp mRNA EST 18-MAY-2000
 LOCUS AW838955
 DEFINITION CM2-LT0061-180200-094-h05 LT0061 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW838955
 VERSION AW838955.1 GI:7932929
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 393)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

TITLE
JOURNAL
MEDLINE
COMMENT

'M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=st2-cm2-LT0061-180
200-094-h054t3=2000-02-18&L4=1)
Seq primer: puc 18 forward
High quality sequence stop: 393.

FEATURES

source
1. 393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LT0061"
/dev_stage="Adult"
/note="Organ: telomios; Vector: puc18; Site.1: SmaI;
Site.2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 90 a 99 c 115 g 89 t
ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1780 tactattgactctctga 1799
|||||
DB 68 TACTATTGACTCTCTTGA 49

RESULT 45
AA294540 415 bp mRNA EST 12-NOV-1997
LOCUS SMOV3MCA761SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SI96MLW-Ovml3) Onchocerca volvulus cDNA clone SMOV3CO761 5', mRNA
sequence.
ACCESSION AA294540
VERSION AA294540.1 GI:2099534
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 415)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: bluescript SK.

FEATURES
Location/Qualifiers

source

1. 415
/organism="Onchocerca volvulus"
/strain="kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMV3CO761"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SI96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-Zap XR; Site.1: Eco RI; Site.2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10⁶ independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigman@bc.org)."
BASE COUNT 110 a 89 c 103 g 102 t 11 others
ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2329 gaagaagatgcataaact 2348
|||||
DB 17 GAAGAAGATGCTATTAACCT 36

RESULT 46
T81898 415 bp mRNA EST 15-MAR-1995
LOCUS yd29a11.s1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone
DEFINITION IMAGE:109628 3', mRNA sequence.
ACCESSION T81898
VERSION T81898.1 GI:704905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Other_ESTs: yd29a11.r1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 714
High quality sequence stops: 314 Source: IMAGE Consortium, LML
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 714 Std Error: 0.00
Seq primer: T3
High quality sequence stop: 314.

FEATURES
Location/Qualifiers

```

source
1. .415
/organism="Homo sapiens"
/db_xref="GDB:465245"
/db_xref="taxon:9606"
/clone="IMAGE:109628"
/clone_1b="Soares fetal liver spleen 1NFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pRTT3D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGCAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pRTT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Felina Bonaldo."

BASE COUNT      85 a      126 c      108 g      88 t      8 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Length 415;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 613 ctccagcagggccagggctgg 632
|||||
Db 133 CTCAGCAGGCCCGCCAGGGTGG 114

RESULT 47
AA294143 439 bp mRNA EST 12-NOV-1997
LOCUS AA294143
DEFINITION SMOV3MCA1534SK Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3) Onchocerca volvulus cDNA SML3CO1534 5', mRNA
sequence.
AA294143
VERSION AA294143.1 GI:2099137
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 439)
AUTHORS Williams,S.A., Licotte-Maniewski,M., Laney,S. and Lustigman,S.
JOURNAL Genes expressed in molting L3 larvae of Onchocerca volvulus
COMMENT Unpublished (1997)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pbuescript SK.
Location/Qualifiers
1. .439
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SML3CO1534"
/clone_1b="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/notes="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by

```

```

day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Licotte-Maniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigmen@ncpc.org)."

BASE COUNT      112 a      91 c      107 g      123 t      6 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Length 439;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 gaagaagatgctataaact 2348
|||||
Db 42 GAAGAAGATGCTATAAAGCT 61

RESULT 48
AQ209355 483 bp DNA GSS 18-SEP-1998
LOCUS AQ209355
DEFINITION HS_3240_A2_B12_T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3240 Col-24 Row-C, DNA sequence.
ACCESSION AQ209355
VERSION AQ209355.1 GI:3622090
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 483)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3240 row: C column: 24
Class: BAC ends
High quality sequence stop: 483.
Location/Qualifiers
1. .483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT      137 a      135 c      102 g      108 t      1 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Length 483;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 aattgtcagtgctctgga 1829

```



```

Db      324 AATGTGCAAGTGTCTGCA 343
|||||
RESULT  49
LOCUS   BF601894/c
DEFINITION 266990 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF601894
VERSION   BF601894.1 GI:11699117
KEYWORDS EST.
SOURCE   cow.
ORGANISM Bos taurus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
           Bovidae; Bovinae; Bos.
REFERENCE
  1 (bases 1 to 491)
  Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
  Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
  ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
  Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
  Keeler,J.W.
  Sequence evaluation of four pooled-tissue normalized bovine cDNA
  libraries and construction of a gene index for cattle
  Genome Res. 11 (4), 626-630 (2001)
  21180013
COMMENT
  Contact: Smith TPL
  USDA, ARS, US Meat Animal Research Center
  PO Box 166, Clay Center, NE 68933-0166, USA
  Tel: 402 762 4366
  Fax: 402 762 4390
  Email: smith@email.marc.usda.gov
  Single pass sequencing. Bases called and alt_trimmed with phred
  v0.980904.e. Vector identified by cross_match with the -minscore 18
  and -mismatch 12 options.
  PCR Primers
  FORWARD: AGGAACAGCTATGACCAT
  BACKWARD: GTTTCACAGTCACGAGC
  Plate: 42 row: N column: 11
  Seq primer: ATTTAGTGACACTATAG.
FEATURES
  source
    Location/Qualifiers
      1..491
        /organism="Bos taurus"
        /db_xref="taxon:9913"
        /clone_lib="MARC 3BOV"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
        library made from pooled tissue from marrow, alveolar
        macrophage, ovary, fetal semitendinosus muscle, and fetal
        longissimus muscle."
BASE COUNT 120 a 104 c 105 g 162 t
ORIGIN
Query Match 0.7%; Score 20; DB 11; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2752 agactcacagatcacagat 2771
|||||
Db 284 AGACTCACAGATCACAGAT 265

```

```

ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 498)
  NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
JOURNAL
  Contact: Robert Strussberg, Ph.D.
  Email: c9apbs-remail.nih.gov
  This clone is available royalty-free through LIND; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:508536
COMMENT
  This read is a RESSEQUENCE of a previously sequenced mouse clone
  This read has been verified (found to hit its original self in the
  correct orientation)
  Seq primer: -40RP from Glibco
  High quality sequence stop: 484.
FEATURES
  source
    Location/Qualifiers
      1..498
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="IMAGE:864448"
        /clone_lib="Soares_mammary_gland_NbMNG"
        /sex="male"
        /tissue_type="mammary gland"
        /dev_stage="4 weeks"
        /lab_host="DH10B"
        /note="Organ: mammary gland; Vector: pT73D-Pac (pharmacia
        RI; 1st strand cDNA was primed with a Not I - oligo(dT)
        primer [5'
        TGTATCCATCTGAGTGGAGCGCGCCGCAATGCTTTTGTGTGTGTGTGTGTGTGT
        T 3']; double-stranded cDNA was ligated to Eco RI
        adaptors (Pharmacia), digested with Not I and cloned into
        the Not I and Eco RI sites of the modified pT73 vector.
        RNA provided by Dr. Minoru Ko, Wayne State Univ. library
        constructed and normalized by Bento Soares and M. Fatima
        Bonalido."
BASE COUNT 138 a 130 c 108 g 121 t 1 others
ORIGIN
Query Match 0.7%; Score 20; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1292 cagctcaaggttcaagcca 1311
|||||
Db 498 CAGCTCAAGGTTCAAGCCA 479

```

```

RESULT  51
LOCUS   BF602580
DEFINITION 267845 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF602580
VERSION   BF602580.1 GI:11699804
KEYWORDS EST.
SOURCE   cow.
ORGANISM Bos taurus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
           Bovidae; Bovinae; Bos.
REFERENCE
  1 (bases 1 to 505)
  Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
  Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
  ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
  Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
  Keeler,J.W.
  Sequence evaluation of four pooled-tissue normalized bovine cDNA
  libraries and construction of a gene index for cattle
  Genome Res. 11 (4), 626-630 (2001)
  21180013
COMMENT
  Contact: Smith TPL
  USDA, ARS, US Meat Animal Research Center
  PO Box 166, Clay Center, NE 68933-0166, USA
  Tel: 402 762 4366
  Fax: 402 762 4390
  Email: smith@email.marc.usda.gov
  Single pass sequencing. Bases called and alt_trimmed with phred
  v0.980904.e. Vector identified by cross_match with the -minscore 18
  and -mismatch 12 options.
  PCR Primers
  FORWARD: AGGAACAGCTATGACCAT
  BACKWARD: GTTTCACAGTCACGAGC
  Plate: 42 row: N column: 11
  Seq primer: ATTTAGTGACACTATAG.
FEATURES
  source
    Location/Qualifiers
      1..491
        /organism="Bos taurus"
        /db_xref="taxon:9913"
        /clone_lib="MARC 3BOV"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
        library made from pooled tissue from marrow, alveolar
        macrophage, ovary, fetal semitendinosus muscle, and fetal
        longissimus muscle."
BASE COUNT 120 a 104 c 105 g 162 t
ORIGIN
Query Match 0.7%; Score 20; DB 11; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2752 agactcacagatcacagat 2771
|||||
Db 284 AGACTCACAGATCACAGAT 265

```

```

RESULT  51
LOCUS   BF602580
DEFINITION 267845 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF602580
VERSION   BF602580.1 GI:11699804
KEYWORDS EST.
SOURCE   cow.
ORGANISM Bos taurus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
           Bovidae; Bovinae; Bos.
REFERENCE
  1 (bases 1 to 505)
  Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
  Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
  ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
  Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
  Keeler,J.W.
  Sequence evaluation of four pooled-tissue normalized bovine cDNA
  libraries and construction of a gene index for cattle
  Genome Res. 11 (4), 626-630 (2001)
  21180013
COMMENT
  Contact: Smith TPL
  USDA, ARS, US Meat Animal Research Center
  PO Box 166, Clay Center, NE 68933-0166, USA
  Tel: 402 762 4366
  Fax: 402 762 4390
  Email: smith@email.marc.usda.gov
  Single pass sequencing. Bases called and alt_trimmed with phred
  v0.980904.e. Vector identified by cross_match with the -minscore 18
  and -mismatch 12 options.
  PCR Primers
  FORWARD: AGGAACAGCTATGACCAT
  BACKWARD: GTTTCACAGTCACGAGC
  Plate: 42 row: N column: 11
  Seq primer: ATTTAGTGACACTATAG.
FEATURES
  source
    Location/Qualifiers
      1..491
        /organism="Bos taurus"
        /db_xref="taxon:9913"
        /clone_lib="MARC 3BOV"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
        library made from pooled tissue from marrow, alveolar
        macrophage, ovary, fetal semitendinosus muscle, and fetal
        longissimus muscle."
BASE COUNT 120 a 104 c 105 g 162 t
ORIGIN
Query Match 0.7%; Score 20; DB 11; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2752 agactcacagatcacagat 2771
|||||
Db 284 AGACTCACAGATCACAGAT 265

```

JOURNAL MEDLINE
 COMMENT
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.
 PCR primers
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCGCCAGTCAGCAGC
 Plate: 44 row: K column: 14
 Seq primer: ATTAGTGACACTATAG.
 location/Qualifiers
 1..505
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_1lb="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."
 155 c 132 g 122 t

BASE COUNT
 ORIGIN
 96 a 155 c 132 g 122 t

Query Match 0.7%; Score 20; DB 11; Length 505;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 tggccatgctgctgaactg 717
 ||||||||||||||||||
 Db 174 TGCCATCTCTGTAAGCTG 193

RESULT 52
 BG382487/c 519 bp mRNA EST 12-MAR-2001
 LOCUS
 DEFINITION
 298398 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION
 BG382487
 VERSION
 BG382487.1 GI:13306959
 KEYWORDS
 EST.
 SOURCE
 pig.
 ORGANISM
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 519)
 Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
 and Keele,J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.
 PCR primers
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCGCCAGTCAGCAGC
 Plate: 4 row: B column: 11
 Seq primer: ATTAGTGACACTATAG.
 location/Qualifiers

JOURNAL COMMENT
 TITLE
 DESIGN AND USE OF TWO POOLED TISSUE NORMALIZED CDNA LIBRARIES FOR
 EST DISCOVERY IN SWINE
 UNPUBLISHED (2000)
 CONTACT: SMITH TPL
 USDA, ARS, US MEAT ANIMAL RESEARCH CENTER
 PO BOX 166, CLAY CENTER, NE 68933-0166, USA
 TEL: 402 762 4366
 FAX: 402 762 4390
 EMAIL: SMITH@EMAIL.MARC.USDA.GOV
 SINGLE PASS SEQUENCING. BASES CALLED AND ALT-TRIMMED WITH PHRED
 V0.980904.E. VECTOR IDENTIFIED BY CROSS_MATCH WITH THE -MINSCORE 18
 AND -MISMATCH 12 OPTIONS.
 PCR PRIMERS
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCGCCAGTCAGCAGC
 PLATE: 4 ROW: B COLUMN: 11
 SEQ PRIMER: ATTAGTGACACTATAG.
 LOCATION/QUALIFIERS

source
 1..519
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_1lb="MARC 1P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."
 103 c 104 g 162 t

BASE COUNT
 ORIGIN
 150 a 103 c 104 g 162 t

Query Match 0.7%; Score 20; DB 11; Length 519;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1788 tgactcttgaacattgc 1807
 ||||||||||||||||||
 Db 312 TGACTTCTTGACACTTGC 293

RESULT 53
 A1670486 521 bp mRNA EST 17-MAY-1999
 LOCUS
 DEFINITION
 SMOV3MCA26G12SK Onchocerca volvulus molting L3 larva cDNA
 (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA26G12 5',
 mRNA sequence.
 A1670486
 VERSION
 A1670486.1 GI:4836992
 KEYWORDS
 EST.
 SOURCE
 Onchocerca volvulus.
 ORGANISM
 Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 1 (bases 1 to 521)
 Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
 Genes expressed in molting L3 larvae of Onchocerca volvulus
 Unpublished (1997)
 JOURNAL
 COMMENT
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genomemail@smith.edu
 Seq primer: phnuescript SK.
 location/Qualifiers
 1..521
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone_1lb="SMOV3MCA26G12"
 /clone_1lb="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLW-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XJ1-Blue MRF"
 /note="Vector: lambda uni-zap XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molt to fourth-stage larvae by
 day 5 in culture. mRNA was isolated from approximately
 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dT) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 uni-zap XR vector and has 1 x 10⁶ independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email: slustigmen@bc.org)."

BASE COUNT 139 a 105 c 125 g 147 t 5 others

ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2329 gaagaagatgctataaact 2348
|||||
Db 42 GAAGAGATGCTATTAACCT 61

RESULT 54
FR0031506 521 bp DNA GSS 27-JUN-1998
LOCUS Fugu rubripes GSS sequence, clone 116B21aA4, genomic survey
DEFINITION

ACCESSION AL027875.1 GI:3269989
KEYWORDS GSS: genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes

REFERENCE
AUTHORS Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranta, Y.,
Williams, G., and Brenner, S.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
biohelp@hmp.mrc.ac.uk
Vector: pBluescript II KS
V.Type: phagemid
PRIMER: KS

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source location/Qualifiers
1..521

/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_id="cosmid 116B21"
/clone="116B21aA4"

BASE COUNT 125 a 120 c 120 g 136 t 20 others
ORIGIN

Query Match 0.7%; Score 20; DB 13; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2585 aagatggaatgaagctctt 2604
|||||
Db 312 AAGATGAATGAAGCTCTT 293

RESULT 55
B1394563 526 bp mRNA EST 06-AUG-2001
LOCUS B1394563
DEFINITION pp1n.pk001.110 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus cDNA clone pp1n.pk001.110 5' similar to
g117710042 ref1NP.057930.11 IQ motif containing GTPase activating
protein 1: cdc42-Rac1 effector protein [Mus musculus]

sp1093KFI10G1.MOUSE RAS GTPASE-ACTIVATING-LIKE PROTEIN IQGAP1
gblAA60344.11(Af240630_1 (Af240630) IQ motif containing GTPas, mRNA
sequence.
B1394563
ACCESSION B1394563
VERSION B1394563.1 GI:15087845

KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 526)

Porter, T.E. and Cogburn, L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IRAFs Animal Genome Project
unpublished (2001)

CONTACT: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES
source location/Qualifiers
1..526

/organism="Gallus gallus"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"

/clone="pp1n.pk001.110"
/clone_id="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (w1,w3,w5
w7,w9)"
/lab_host="E. Coli EMDH10B"

/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"
BASE COUNT 124 a 162 c 151 g 87 t 2 others
ORIGIN

Query Match 0.7%; Score 20; DB 11; Length 526;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 705 gctgctgaagctgcgcaga 724
|||||
Db 72 GCTGCTGAAGCTGC GCACAGA 91

RESULT 56
BG351255 535 bp mRNA EST 01-MAR-2001
LOCUS BG351255
DEFINITION 08BD06 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
sequence.

ACCESSION BG351255
KEYWORDS BG351255.1 GI:13179997
EST.
SOURCE potato.
ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 535)

Nielsen, K.L., Crookshanks, M., Emmersen, J. and Wellinder, K.G.
EST-sequencing of mature potato tuber (Var. Kuras)

unpublished (2000)
CONTACT: Karen G. Wellinder
Institute for biotechnology
Aalborg University

Sohngarsholmsvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808

Email: kgw@bio.auc.dk
Sequenced from the 5' end.
High quality sequence stop: 535
POLYA-No.

```

FEATURES
  source      Location/Qualifiers
              1..535
              /organism="Solanum tuberosum"
              /cultivar="Field grown Kuras"
              /db_xref="taxon:4113"
              /clone_lib="Mature tuber lambda ZAP"
              /tissue_type="Tuber"
              /note="Vector: Lambda ZAP"
BASE COUNT   177 a      110 c      109 g      139 t
ORIGIN
Query Match      0.7%; Score 20; DB 11; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 tgaataaacctgaagaagatg 2379
      |||||||
Db 160 TGAATAAACCTGAAGAAGATG 179

RESULT 57
LOCUS      A2221490      541 bp      DNA      GSS      14-JUN-2000
DEFINITION Gm_UMB001_125_D12.R UMN Soybean BAC library (PECSBAC4 EcORI)
ACCESSION  A2221490
VERSION     A2221490.1 GI:8517274
KEYWORDS
SOURCE      soybean.
            Glycine max
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
            1 (bases 1 to 541)
REFERENCE   1 (bases 1 to 541)
            Marek, L.F., Paz, M., Darnielle, L., Hanson, N. and Shoemaker, R.C.
            BAC End sequences from a soybean genomic library (ISU)
            Unpublished (2000)
            Contact: Shoemaker Randy C
            Agronomy Department
            Iowa State University
            Ames, IA 50011-1010, USA
            Tel: 515 294 1205
            Fax: 515 294 2299
            Email: rcschoe@iastate.edu
            This BAC identified by SSR Satt427. For more information, see
            Soybase at:
            http://genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase.
            Please see as an authority for the mapping/naming: Cregan P.B., T.
            Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Larik, A.L. Kahler, N. Kaya,
            T.T. Vantolai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An
            integrated genetic linkage map of the soybean genome. Crop Sci.
            39:(in press)
            Class: BAC ends.
            Seq primer: BACR or M13R
FEATURES
  source      Location/Qualifiers
              1..541
              /organism="Glycine max"
              /cultivar="Fairbault"
              /db_xref="taxon:3847"
              /clone_lib="Gm_UMB001_125_D12"
              /clone_lib="UMN Soybean BAC Library (PECSBAC4 EcORI)"
              /tissue_type="cotyledon leaves"
              /dev_stage="cotyledon stage"
              /note="Vector: PECSBAC4; The UMN BAC library (Danesh et al
              , Theor. Appl. Genet. 96:196, 1998) was constructed using
              the Eco RI site of PECSBAC4. The library consists of 72
              ,960 clones with an average insert size of 120 Kb, equal
              to 7 haploid genome equivalents. Screening of the library
              is done by PCR amplification of DNA pools."
BASE COUNT   177 a      88 c      90 g      165 t      21 others
ORIGIN

```

```

Query Match      0.7%; Score 20; DB 13; Length 541;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 989 gcttgagaatcctcatgaag 1008
      |||||||
Db 424 GCTTGAGGAATCTCATGAG 443

RESULT 58
LOCUS      A2298434      554 bp      DNA      GSS      27-JUL-2000
DEFINITION RPCI-23-109A14.TV RPCI-23 Mus musculus genomic clone RPCI-23-109A14
            , DNA sequence.
ACCESSION  A2298434
VERSION     A2298434.1 GI:9540219
KEYWORDS
SOURCE      house mouse.
            Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 554)
REFERENCE   1 (bases 1 to 554)
            Zhao, S., Mierman, W., Feldblum, T., Malek, J., Shatsman, S., Akiret
            , B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
            and Frazer, C.M.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
            Other_GSSs: RPCI-23-109A14.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.bufileo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/cdb/bac-ends/mouse/bac-end_intro.html
            Plate: 109 row: A column: 14
            Seq primer: 17
            Class: BAC ends.
FEATURES
  source      Location/Qualifiers
              1..554
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone_lib="RPCI-23-109A14"
              /clone_lib="RPCI-23"
              /sex="Female"
              /lab_host="DH10B"
              /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;
              EcORI; Site: 2; EcORI; Female C57BL/6J mouse kidney and/or
              brain genomic DNA was isolated and partially digested
              with a combination of EcORI and EcORI Methylase. Size
              selected DNA was cloned into the pBACe3.6 vector at the
              EcORI sites. The ligation products were transformed into
              DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT   179 a      80 c      76 g      219 t
ORIGIN
Query Match      0.7%; Score 20; DB 13; Length 554;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1833 cattaactgactcttatg 1852
      |||||||
Db 97 CATTAACCTGACTCTTATG 116

```

RESULT 59
LOCUS AQ270247/c
DEFINITION HS_2045_AL_G04_MK_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2045 Col=7 Row=M, DNA sequence.
ACCESSION AQ270247
VERSION AQ270247.1 GI:3822842
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 557)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2045 row: M column: 7
Class: BAC ends
High quality sequence stop: 557.
FEATURES
source location/Qualifiers
1..557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2045 Col=7 Row=M"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in E-Coli DH10B"

BASE COUNT 137 a 140 c 108 g 169 t 3 others
ORIGIN

Query Match 0.7%; Score 20; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1905 catgaagagcccccagaa 1924
|||||
Db 192 CATGAGAGCGCCCGAGAA 173

RESULT 60
LOCUS BF258009/c
DEFINITION BF258009 571 bp mRNA EST 23-FEB-2001
HVSMEF0014H19f Hordeum vulgare seedling root EST library HVCDNA0007 (etioloated and unstressed) Hordeum vulgare cDNA clone
HVSMEF0014H19f, mRNA sequence.
ACCESSION BF258009
VERSION BF258009
KEYWORDS BF258009.2 GI:13119185
SOURCE EST.
ORGANISM barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 571)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
T., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and

TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11187122.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTACCCCTCCTCAAGG
High quality sequence stop: 558.
FEATURES
source location/Qualifiers
1..571
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEF0014H19f"
/clone_lib="Hordeum vulgare seedling root EST library HVCDNA0007 (etioloated and unstressed)"
/issue_type="Seedling root"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI; For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley/> to order a clone see <http://www.genome.clemson.edu/orders>"

BASE COUNT 101 a 166 c 218 g 86 t
ORIGIN

Query Match 0.7%; Score 20; DB 11; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 465 tggcctctgcagctcttc 484
|||||
Db 191 TGGCTCTCTGCAGCTCTTC 172

RESULT 61
LOCUS BG385206/c
DEFINITION BG385206 577 bp mRNA EST 12-MAR-2001
306845 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BG385206
VERSION BG385206.1 GI:13309678
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 577)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 94 row: J column: 10
Seq primer: ATTAGGTGACACTATAG.

FEATURES

source

Location/Qualifiers

1. 577
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 1P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."
 BASE COUNT 102 a 200 c 175 g 100 t

ORIGIN

Query Match 0.7%; Score 20; DB 11; Length 577;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 385 aacttgaaagcaccctc 404
 ||||||||||||||||
 Db 572 AACTGMAAGCACCTTCAC 553

RESULT 62

A0570315

577 bp DNA

GSS

01-JUN-1999

LOCUS

HS_5352_B1_E02_T7A RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=928 Col=3 Row=J, DNA sequence.

ACCESSION

A0570315

VERSION

A0570315.1

KEYWORDS

GI:4963535

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 577)

AUTHORS

Mahaltras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE

99380589

COMMENT

Contact: Mahaltras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pliet de Jong
 (pliederde@med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 928 Row: J Column: 3
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 577.

FEATURES

source

Location/Qualifiers

1. 577

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate=928 Col=3 Row=J"

/clone_lib="RPCI-11 Human Male BAC Library"

/sex="Male"

/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated with a combination of EcoRI and
 partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"

BASE COUNT

244 a 114 c 82 g 130 t 7 others

ORIGIN

Query Match 0.7%; Score 20; DB 13; Length 577;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2070 cagctgcgaataaagat 2089
 ||||||||||||||||
 Db 348 CAGGCTGCAATTAAGAGAT 367

RESULT 63

A2987277/c

580 bp DNA

GSS

27-APR-2001

LOCUS

2M0269J15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0269J15 R, DNA sequence.

DEFINITION

A2987277

ACCESSION

A2987277.1

VERSION

GI:13858504

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0265 Row: J Column: 15
 Seq primer: CACACAGAAACGCTATGACC
 Class: Plasmid ends
 High quality sequence stop: 580.

FEATURES

source

Location/Qualifiers

1. 580

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0269J15"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, M1-resistant, F-"
 /note="Vector: pMD42ny; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (911473211419b/AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

```

BASE COUNT      169 a      91 c      93 g      227 t
ORIGIN

Query Match      0.7%; Score 20; DB 13; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1303 ttcaagccaagataaatt 1322
|||||
Db 479 TTCAGCCCAAGTATTAAT 460

RESULT 64
BC098273      605 bp      mRNA      EST      29-JAN-2001
LOCUS      EST1462792 sprouting eyes/shoots Solanum tuberosum cDNA clone
DEFINITION      cSTC2015 5' sequence, mRNA sequence.
ACCESSION      BC098273
VERSION      BC098273.1 GI:12588308
KEYWORDS      EST.
SOURCE      potato.
ORGANISM      Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 605)
van der Hoeven,R.S., Bezzerides,J., Cho,J., Utterback,T., Hansen
,C.L., Bougri,O., Beull,C.R., Romning,C.M., Tanksley,S.D. and Baker
,B.
Generation of ESTs from potato sprouting eyes/shoots
Unpublished (2001)
Contact: Cathy Romning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@esgen.com.
location/Qualifiers
source
1. 605
/organism="Solanum tuberosum"
/cultivar="kennebec"
/db_xref="taxon:4113"
/clone="cSTC2015"
/clone_lib="sprouting eyes/shoots"
/tissue_type="sprouting tubers"
/seq_strategy="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2 mm to 15 mm) were
taken from tubers. The tubers were incubated at 26C in
the dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."

BASE COUNT      192 a      117 c      119 g      177 t
ORIGIN

Query Match      0.7%; Score 20; DB 11; Length 605;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2360 tgaaaacctgaagaagatg 2379
|||||
Db 156 TGA AAAACCTGAGAGATG 175

RESULT 65
A2802117      620 bp      DNA      GSS      16-FEB-2001
LOCUS      2M0061H04F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
DEFINITION      clone UNGC2M0061H04 F, DNA sequence.
ACCESSION      A2802117
VERSION      A2802117.1 GI:12954536
KEYWORDS      GSS.

```

```

SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 620)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0061 row: H column: 04
Seq primer: CGTGTAAACGACGCCACGT
Class: plasmid ends
High quality sequence stop: 620.
location/Qualifiers
source
1. 620
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UNG2M0061H04"
/clone_lib="Mouse 10kb plasmid UNGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g11473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      167 a      141 c      175 g      137 t
ORIGIN

Query Match      0.7%; Score 20; DB 13; Length 620;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2110 ctcaattgtctccagac 2129
|||||
Db 180 CTCAGTTGTCTCAGCAC 199

RESULT 66
AT511575      627 bp      mRNA      EST      16-MAR-1999
LOCUS      SMOVL3CAN30H03K Onchocerca volvulus infective larva cDNA
DEFINITION      (SAB94WL-OVL3) Onchocerca volvulus cDNA clone SMOVL3CAN30H03 5',
mRNA sequence.
ACCESSION      AT511575

```

VERSION AI511575.1 GI:4417429
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 627)
AUTHORS Williams,S.A., Lu,W., Lizotte-Maniewski,M. and Laney,S.J.
TITLE Genes expressed in infective third stage larvae of Onchocerca volvulus
JOURNAL Unpublished (1995)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
1. 627
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone_lib="SMOVL3CANCAN30H03"
/clone_lib="Onchocerca volvulus infective larva cDNA (SAM94WL-OVL3)"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Unizap XR; Site_1: EcoR I; Site_2: Xho I; Cutaneous filarial nematode parasite of humans. mRNA was prepared from third stage infective larvae of Onchocerca volvulus isolated from mosquitoes 10 days after infection and converted to double stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNase I. The library had 1.8 x 10E5 independent recombinants and average insert size was 900 base pairs. The library was constructed by Wenhong Lu. The library is available from Dr. S.A. Williams, email genome@smith.edu."
BASE COUNT 173 a 123 c 148 g 183 t
ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 gaagaagatgcataaact 2348
|||||
Db 28 GAAGAGAGTCTATAAACT 47

RESULT 67
A2510047 633 bp DNA GSS 05-OCT-2000
LOCUS IM0354H14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0354H14 F, DNA sequence.
ACCESSION A2510047
VERSION A2510047.1 GI:10691363
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 633) Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL plasmid inserts
COMMENT Unpublished (2000)
Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0354 row: H column: 14
Seq primer: CGTGTAAACGACGCCGCGT
Class: plasmid ends
High quality sequence stop: 633.
Location/Qualifiers
1. 633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0354H14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XLI0-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114[gb]/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XLI0-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 147 a 154 c 134 g 198 t
ORIGIN

Query Match 0.7%; Score 20; DB 13; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 tggcacaatcagaagcaga 691
|||||
Db 257 TGGCACATCAGAGAGCAGA 276

RESULT 68
AM874886 644 bp mRNA EST 22-MAY-2000
LOCUS SMOVFCAR10G04SK Onchocerca volvulus microfilaria cDNA
DEFINITION (SAM98MLM-OVMf) Onchocerca volvulus cDNA clone SMOVFCAR10G04 5',
mRNA sequence.
ACCESSION AM874886
VERSION AM874886.1 GI:8012581
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 644)
AUTHORS Williams,S.A.
TITLE Genes expressed in microfilaria of Onchocerca volvulus
JOURNAL Unpublished (1999)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA

Tel: 4135853826
Fax: 4135853786

Email: genome@smith.edu
Seq primer: Bluescript SK.

Location/Qualifiers

1. .644

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone="SMOV3MCAM07B09" /clone="SMOV3MCAM07B09"

(SAW98MLM-OVME)"

/dev_stage="microfilaria"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. mRNA was
prepared from approximately 200,000 microfilariae isolated
from the skin of infected individuals from Kumba,
Cameroon and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library has 7.8 x 10E4 independent
recombinants and the average insert size is approximately
1kb. The library was constructed by Michelle
Lizotte-Waniewski. The library is available from
Dr.S.A.Williams, email:genome@smith.edu."

BASE COUNT 182 a 125 c 152 g 185 t

ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 644;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2329 gaagaagatgcataaact 2348

Db 27 GAAGAAGATGCTATAAACT 46

RESULT 69
BE346038

LOCUS BE346038 652 bp mRNA EST 17-JUL-2000
DEFINITION SMOV3MCAM07B09SK Onchocerca volvulus L3 larvae cDNA (SAW98MLM-OVME2)
ACCESSION BE346038
VERSION BE346038
KEYWORDS BE346038.1 GI:9255574

EST.

ORGANISM Onchocerca volvulus.

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 652)

AUTHORS Williams, S.A.

TITLE Genes expressed in L2 larvae of Onchocerca volvulus

JOURNAL Unpublished (1999)

COMMENT Contact: Steven A. Williams

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu

Seq primer: Bluescript SK.

Location/Qualifiers

1. .652

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone="SMOV3MCAM07B09" /clone="SMOV3MCAM07B09"

(SAW98MLM-OVME2)"

/dev_stage="L2" /dev_stage="L2"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. mRNA was
prepared from approximately 9,000 L2s isolated from
infected mosquitoes from Kumba, Cameroon and converted to
double-stranded cDNA using reverse transcriptase and
oligo(dT) followed by RNase H and DNA pol I. The library
has 7.3 x 10E4 independent recombinants and the average
insert size is approximately 1kb. The library was
constructed by Michelle Lizotte-Waniewski. The library is
available from Dr.S.A.Williams, email: genome@smith.edu."

BASE COUNT 187 a 124 c 157 g 184 t

ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2329 gaagaagatgcataaact 2348

Db 27 GAAGAAGATGCTATAAACT 46

RESULT 70
A1317901

LOCUS A1317901 653 bp mRNA EST 17-DEC-1998
DEFINITION SMOV3MCAM07B09SK Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-OVMEJ3) Onchocerca volvulus cDNA clone SMOV3MCAM07B09 5',
mRNA sequence.

ACCESSION A1317901
VERSION A1317901
KEYWORDS A1317901.1 GI:4033168

EST.

ORGANISM Onchocerca volvulus.

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 653)

AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.

TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus

JOURNAL Unpublished (1997)

COMMENT Contact: Steven A. Williams

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu

Seq primer: Bluescript SK.

Location/Qualifiers

1. .653

/organism="Onchocerca volvulus"

/strain="Kumba, Cameroons"

/db_xref="taxon:6282"

/clone="SMOV3MCAM07B09" /clone="SMOV3MCAM07B09"

(SL96MLM-OVMEJ3)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H

and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent

recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle

lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nhydc.org)."

BASE COUNT 184 a 127 c 160 g 182 t

ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 653;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 gaagaagatgctataaact 2348
|||||

DB 31 GAAGAAGATGCTATAAACT 50

RESULT 71
LOCUS BE636495 656 bp mRNA EST 25-AUG-2000
DEFINITION SMOVL2CASI7D04SK Onchocerca volvulus L2 larvae cDNA (SAM98MLM-OVL2)
ACCESSION BE636495
VERSION BE636495.1 GI:9919606
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
REFERENCE 1 (bases 1 to 656)
AUTHORS Williams,S.A.
TITLE Genes expressed in L2 larvae of Onchocerca volvulus
JOURNAL Unpublished (1999)
COMMENT Molecular Parasitology
Smith College Department of Biological Sciences
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1..656
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOVL2CASI7D04"
/clone.lib="Onchocerca volvulus L2 larvae cDNA (SAM98MLM-OVL2)"
/dev_stage="L2"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 9,000 L2s isolated from infected mosquitoes from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.3 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr. S.A. Williams, email: genome@smith.edu."

BASE COUNT 183 a 129 c 157 g 186 t

ORIGIN

Query Match 0.7%; Score 20; DB 10; Length 656;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 gaagaagatgctataaact 2348
|||||

DB 28 GAAGAAGATGCTATAAACT 47

RESULT 72
LOCUS BG310513 660 bp mRNA EST 23-FEB-2001
DEFINITION SMOV3MCA56C02SK Onchocerca volvulus molting L3 larva cDNA (SL96MLM-OVml3) Onchocerca volvulus cDNA clone SMOV3MCA56C02 5', mRNA sequence.

ACCESSION BG310513
VERSION BG310513.1 GI:13112371
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
REFERENCE 1 (bases 1 to 660)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1..660
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCA56C02"
/clone.lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-OVml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCIC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nhydc.org)."

BASE COUNT 187 a 127 c 161 g 185 t

ORIGIN

Query Match 0.7%; Score 20; DB 11; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 gaagaagatgctataaact 2348
|||||

DB 28 GAAGAAGATGCTATAAACT 47

RESULT 73
LOCUS CNS02935/c 726 bp DNA GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence pUC-ori end of clone 247A05 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL186746

VERSION AL186746.1 GI:7824850
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 726)
AUTHORS Ruest-Crolius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNLS Unpublished
REFERENCE 2 (bases 1 to 726)
AUTHORS Ruest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 726)
AUTHORS Genoscope.
TITLE Direct Submission
COMMENT Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..726
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="247A05"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG247AA03SP1-end :
PUC-ori"
BASE COUNT 231 a 162 c 139 g 185 t 9 others
ORIGIN
Query Match 0.7%; Score 20; DB 13; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 180 aaaggttcagagtcctgta 199
|||||
Db 533 AAAGGTCACAGCTCCTGA 514
RESULT 74
B65395 729 bp DNA GSS 21-JUN-1998
LOCUS CIT-HSP-2021M16.TF CIT-HSP Homo sapiens genomic clone 2021M16, DNA
DEFINITION sequence.
ACCESSION B65395
VERSION B65395.1 GI:2639373
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 729)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Contact: Mark Adams
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..729
/organism="Homo sapiens"
/db_xref="GDB:7045554"
/db_xref="taxon:9606"
/clone="2021M16"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11; site_1: HindIII; site_2:
HindIII"
BASE COUNT 241 a 134 c 126 g 228 t
ORIGIN
Query Match 0.7%; Score 20; DB 13; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2343 aaactagctgaagggcctga 2362
|||||
Db 673 AAACCTAGCTGAGGCCCTGA 654
RESULT 75
BG250425 731 bp mRNA EST 13-FEB-2001
LOCUS BG250425
DEFINITION 602362580P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4471147 5',
mRNA sequence.
ACCESSION BG250425
VERSION BG250425.1 GI:12760241
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 731)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L1AM10289 row: 0 column: 20
High quality sequence stop: 611.
FEATURES
source
Location/Qualifiers
1..731
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4471147"
/clone_1lb="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; site_1: NotI;
site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 215 a 185 c 165 g 166 t
ORIGIN

Query Match 0.7%: Score 20; DB 11; Length 731;
Best Local Similarity 100.0%: Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1114 cagaaacacaacaaca 1133
 |||||
Db 658 CAGAAAACACACAACA 677

Search completed: March 25, 2002, 15:28:33
Job time: 13682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 16:50:28 ; Search time 270.41 Seconds

(without alignments)
9739.666 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcataagagcaaa.....ctttaactagtaactgct 3072

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

N.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3072	100.0	3133	22 AAS03945 Human caspase recr
2	2919	95.0	3545	22 AAH88254 Murine EST-derived
3	2634	85.7	3615	22 AAS03946 Human caspase recr
4	1781	58.0	2950	22 AAH9581 Human protein enco
c	489	15.9	2735	22 AAH34171 Human colon cancer
5	442	7.9	421	22 AAH34189 Probe #4322 for ge
6	242	7.9	421	22 AAH35764 Probe #4450 used t
7	242	7.9	421	22 AAH35764 Probe #4204 used t
8	220	7.2	220	22 AAH13350 Probe #13523 for g
9	220	7.2	220	22 AAH13350 Probe #17590 used
10	220	7.2	220	22 AAH18904 Probe #17590 used
11	220	7.2	220	22 AAH09206 Probe #9197 used t

12	19	0.6	549	22 AAH11452 Human cDNA clone (
13	19	0.6	579	19 AA330628 H. pylori cell env
c	14	0.6	649	19 AA314445 H. pylori GHP 875
c	15	0.6	1908	21 AAF13098 Aspergillus oryzae
c	16	0.6	2825	22 AAH16202 Human cDNA sequenc
c	17	0.6	116	21 AAC14986 Human secreted pro
c	18	0.6	274	21 AAH31261 Plant microsatelli
c	19	0.6	322	16 AAH24250 Human gene signatu
c	20	0.6	355	21 AAH31416 Plant microsatelli
c	21	0.6	377	21 AAH31325 Plant microsatelli
c	22	0.6	435	21 AAH31366 Plant microsatelli
c	23	0.6	446	21 AAC09333 Human secreted pro
c	24	0.6	454	22 AAH14905 Probe #4838 for ge
c	25	0.6	454	22 AAH36257 Probe #4943 used t
c	26	0.6	454	22 AAH0685 Probe #4676 used t
c	27	0.6	466	21 AAC01809 Human secreted pro
c	28	0.6	468	21 AAH31287 Plant microsatelli
c	29	0.6	472	22 AAH11787 Probe #1720 for ge
c	30	0.6	472	22 AAH33101 Probe #1787 used t
c	31	0.6	472	22 AAH01718 Probe #1709 used t
c	32	0.6	594	11 AAH05868 Sequence encoding
c	33	0.6	652	21 AAH05868 Fusarium venenatum
c	34	0.6	682	21 AAH07389 Arabidopsis thalia
c	35	0.6	814	22 AAH05406 Human cDNA clone (
c	36	0.6	888	20 AAH20212 Enterococcus faeca
c	37	0.6	960	22 AAH67086 C glutamicum codin
c	38	0.6	992	11 AAH05870 Sequence encoding
c	39	0.6	1034	21 AAH25257 Human secreted pro
c	40	0.6	1083	22 AAH71387 Corynebacterium q1
c	41	0.6	1205	20 AAH24236 Human normal blad
c	42	0.6	1250	9 AAH80316 Transcription cont
c	43	0.6	1473	21 AAH40012 Arabidopsis thalia
c	44	0.6	1491	21 AAH47150 DNA encoding a ser
c	45	0.6	1515	19 AAH40737 C. felis esterase,
c	46	0.6	1515	19 AAH40738 C. felis esterase,
c	47	0.6	1561	21 AAH21032 Human low adenosin
c	48	0.6	1561	21 AAH34910 Human adenosine re
c	49	0.6	1579	22 AAH33702 Human cDNA sequenc
c	50	0.6	1611	12 AAH012528 Thymidylate phosph
c	51	0.6	1611	21 AAH47151 DNA encoding a ser
c	52	0.6	1619	21 AAH15908 Human protein clon
c	53	0.6	1650	19 AAH40760 C. felis esterase,
c	54	0.6	1650	19 AAH40760 C. felis esterase,
c	55	0.6	1723	20 AAH40761 C. felis esterase,
c	56	0.6	1800	15 AAH07601 Human vesicular bl
c	57	0.6	1876	22 AAH25374 Retinoblastoma pro
c	58	0.6	1926	20 AAH82522 Nucleotide sequenc
c	59	0.6	1962	20 AAH82522 Sphingomonas capsu
c	60	0.6	1982	19 AAH40735 Human ovarian tumo
c	61	0.6	1982	19 AAH40736 C. felis esterase,
c	62	0.6	2144	19 AAH40758 C. felis esterase,
c	63	0.6	2144	19 AAH40759 C. felis esterase,
c	64	0.6	2299	19 AAH26326 S. pneumoniae deri
c	65	0.6	2299	19 AAH42995 Streptococcus pneu
c	66	0.6	3000	20 AAH84399 S. capsulata JF012
c	67	0.6	3001	21 AAH51776 Human polynucleoti
c	68	0.6	3001	21 AAH51787 Human polynucleoti
c	69	0.6	3275	22 AAH25373 Enterococcus faeca
c	70	0.6	3529	21 AAH31144 HMG-CoA reductase
c	71	0.6	3942	20 AAH80317 Transcription cont
c	72	0.6	3969	21 AAH1374 Human low adenosin
c	73	0.6	4215	22 AAH58415 Human adenosine re
c	74	0.6	4871	22 AAH02020 Human polynucleoti
c	75	0.6	4951	20 AAH31318 Human polynucleoti
c	76	0.6	4982	18 AAH85267 Enterococcus faeca
c	77	0.6	5212	9 AAH80317 Transcription cont
c	78	0.6	5292	21 AAH1374 Human low adenosin
c	79	0.6	5292	21 AAH35252 Human adenosine re
c	80	0.6	5511	17 AAH41853 cDNA encoding Plas
c	81	0.6	6212	21 AAH21375 Human low adenosin
c	82	0.6	6212	21 AAH35253 Human adenosine re
c	83	0.6	7380	20 AAH84028 MMP9 promoter beta
c	84	0.6	7383	22 AAH00153 Matrix metalloprot

|||||
Db 756 cagaaggtctcttctctcttgatgctacaatgtaattcaagccccaagactgccagaa 815
OY 761 atcgaagccctgataaagaagaaacccacgctccaagaacatgctatcgtccactacc 840
Db 816 atcgaagccctgataaagaagaaacccacgctccaagaacatgctatcgtccactacc 875
OY 841 actgagtgctcctgagcaatacagcagtttggtgcccctgacgtcgcgaagtggggatag 900
Db 876 actgagtgctcctgagcaatacagcagtttggtgcccctgacgtcgcgaagtggggatag 935
OY 901 acagaagaacgagcccaagctctccatccgagaagtgtcgtatcaagaagactgtctgaagc 960
Db 936 acagaagaacgagcccaagctctccatccgagaagtgtcgtatcaagaagactgtctgaagc 995
OY 961 ttgtgtccccaattccagaataaccagtgcttgagaatctcataagaagccctctctt 1020
Db 996 ttgtgtccccaattccagaataaccagtgcttgagaatctcataagaagccctctctt 1055
OY 1021 gtggtcactactgtgcacatccagaatgggtggaagtgaagttccactctccacacaaaca 1080
Db 1056 gtggtcactactgtgcacatccagaatgggtggaagtgaagttccactctccacacaaaca 1115
OY 1081 acgctgtccatcactctctatgctgtgtgatacagaagaaacacacaaacataaaggt 1140
Db 1116 acgctgtccatcactctctatgctgtgtgatacagaagaaacacacaaacataaaggt 1175
OY 1141 gtgctgcgaagtgtactcaatccgagagccttgacacactgtggaagcctagctctgagaggt 1200
Db 1176 gtgctgcgaagtgtactcaatccgagagccttgacacactgtggaagcctagctctgagaggt 1235
OY 1201 gtgtctccccaagatttgatctcgaaactgcagagatgtgtccagcgtgaatgaagatgtc 1260
Db 1236 gtgtctccccaagatttgatctcgaaactgcagagatgtgtccagcgtgaatgaagatgtc 1295
OY 1261 ctgctgcgaactgtgctcctctgttaataatacagctcaaaaggtctcaagccaagtataaa 1320
Db 1296 ctgctgcgaactgtgctcctctgttaataatacagctcaaaaggtctcaagccaagtataaa 1355
OY 1321 tctcttcacaagatctctccagagatgacacagcagagacgaagcctcagcagtttatagcg 1380
Db 1356 tctcttcacaagatctctccagagatgacacagcagagacgaagcctcagcagtttatagcg 1415
OY 1381 tctcattgagccagagaggtgagcagaagggaatggttactctgacagaatagtttccatt 1440
Db 1416 tctcattgagccagagaggtgagcagaagggaatggttactctgacagaatagtttccatt 1475
OY 1441 tccgaacattacatccacttatagcagcctgtccggtacacacctgtggtgcattcgttgaa 1500
Db 1476 tccgaacattacatccacttatagcagcctgtccggtacacacctgtggtgcattcgttgaa 1535
OY 1501 gccaccagagcctgttatgaagacactccgagcagtgatatcaacacagcctgtctctcgaa 1560
Db 1536 gccaccagagcctgttatgaagacactccgagcagtgatatcaacacagcctgtctctcgaa 1595
OY 1561 ctttcacatccgaagagcctctctgagagacagagatctctgcaagvtgtgaaaaacac 1620
Db 1596 ctttcacatccgaagagcctctctgagagacagagatctctgcaagvtgtgaaaaacac 1655
OY 1621 actgcgaagaataatctctgaagcgaataaacatcaattccttctgtgaagtgtgtgacatc 1680
Db 1656 actgcgaagaataatctctgaagcgaataaacatcaattccttctgtgaagtgtgtgacatc 1715
OY 1681 ttatatcagaagagatcacatccaaatcacgcccctgagccagaagatttgaagcttcttcaa 1740
Db 1716 ttatatcagaagagatcacatccaaatcacgcccctgagccagaagatttgaagcttcttcaa 1775
OY 1741 ggtaaaagcttatataaactcagaggaaacatcccgaattacttatttgacttctttaa 1800
Db 1776 ggtaaaagcttatataaactcagaggaaacatcccgaattacttatttgacttctttaa 1835
OY 1801 cattgcccaattgtgcgaagtgtcctggaactcattaaactggaacttattatggtggagct 1860

Db 1836 cattgcccaattgtgcgaagtgtcctggaactcattaaactggaacttattatggtggagct 1895
OY 1861 atggtcttaatgggaaaagcctgtgcagaagacagaagtggaatccacatggaagagcccca 1920
Db 1896 atggtcttaatgggaaaagcctgtgcagaagacagaagtggaatccacatggaagagcccca 1955
OY 1921 gaaacctaatccccagaagcctgtatccttgtcttccaactggaagcagaatctcagg 1980
Db 1956 gaaacctaatccccagaagcctgtatccttgtcttccaactggaagcagaatctcagg 2015
OY 1981 actctggaagtgcacatcccggaattccagcaagtgaabaagcaagatatacacatctgt 2040
Db 2016 actctggaagtgcacatcccggaattccagcaagtgaabaagcaagatatacacatctgt 2075
OY 2041 gggaaaataatcagctcgtcccaagcctccagcgtcgaataataagatgtgtgtgtgtg 2100
Db 2076 gggaaaataatcagctcgtcccaagcctccagcgtcgaataataagatgtgtgtgtgtg 2135
OY 2101 gtctgaagcctcagttgtgtccctcagcacctgttaagaacatttatctctcatgttgaa 2160
Db 2136 gtctgaagcctcagttgtgtccctcagcacctgttaagaacatttatctctcatgttgaa 2195
OY 2161 gccagtcctccatccatagaagaatgagagcacaatcacatctgttaacaaacctgaaaac 2220
Db 2196 gccagtcctccatccatagaagaatgagagcacaatcacatctgttaacaaacctgaaaac 2255
OY 2221 ttgaattatcagactctacagaatacaagcgtccgggtgtgtcttactgacagccttggt 2280
Db 2256 ttgaattatcagactctacagaatacaagcgtccgggtgtgtcttactgacagccttggt 2315
OY 2281 aacttgaagaacctacagaagctcataltgaataacataaagatgaagaagatgtc 2340
Db 2316 aacttgaagaacctacagaagctcataltgaataacataaagatgaagaagatgtc 2375
OY 2341 ataaactagctgtgaagccttgaaaaaacctgaaagaatgtgttatttcaatttgaccac 2400
Db 2376 ataaactagctgtgaagccttgaaaaaacctgaaagaatgtgttatttcaatttgaccac 2435
OY 2401 ttgtctgcacattggagagggaaatgtaattacatagtcgaagtctctgtcaagtbaaccgt 2460
Db 2436 ttgtctgcacattggagagggaaatgtaattacatagtcgaagtctctgtcaagtbaaccgt 2495
OY 2461 gaccttgaagaataatcaattagctcctcgtgtcctgtccgcgaagtcagctgaaaaatccta 2520
Db 2496 gaccttgaagaataatcaattagctcctcgtgtcctgtccgcgaagtcagctgaaaaatccta 2555
OY 2521 gctcgaatcttcacaattggttgcacaactgagcattcttgatttatacgaanaattacctgt 2580
Db 2556 gctcgaatcttcacaattggttgcacaactgagcattcttgatttatacgaanaattacctgt 2615
OY 2581 gaaaaagatggaaatggaagctcttcatgaactgatacgaaggaatgaagtgctagaacag 2640
Db 2616 gaaaaagatggaaatggaagctcttcatgaactgatacgaaggaatgaagtgctagaacag 2675
OY 2641 ctccacgcacatgatgtccctcgtgggcgtgtgacgtgcgaagcagcgtcagcagcctgtgtg 2700
Db 2676 ctccacgcacatgatgtccctcgtgggcgtgtgacgtgcgaagcagcgtcagcagcctgtgtg 2735
OY 2701 aaacatttggagaggtccccaactcgtcaaagcttgggtgtgaaaaactggaactcaca 2760
Db 2736 aaacatttggagaggtccccaactcgtcaaagcttgggtgtgaaaaactggaactcaca 2795
OY 2761 gatacagaatagatttgggtgtcatttttttgaaaagaacccctgtgaaaaacttccag 2820
Db 2796 gatacagaatagatttgggtgtcatttttttgaaaagaacccctgtgaaaaacttccag 2855
OY 2821 cagttgaatttggcggaagatcgtgtgacagtgatgatacgtcgtcctcattggtgtga 2880
Db 2856 cagttgaatttggcggaagatcgtgtgacagtgatgatacgtcgtcctcattggtgtga 2915
OY 2881 ttggaatccttaagcaattagtgtttttttgactttagtactaaagaaattctactactgat 2940
Db 2916 ttggaatccttaagcaattagtgtttttttgactttagtactaaagaaattctactactgat 2975

QY 2941 ccagcattagtcagaaacttagccaagtgltatccaagttaactttctgcagaagct 3000
|||||
Db 2976 ccagcattagtcagaaacttagccaagtgltatccaagttaactttctgcagaagct 3035
QY 3001 agcctgtgtgtgtgcgaattgatgatgatccagtgltattacagtgcttttaa 3060
|||||
Db 3036 agcctgtgtgtgtgcgaattgatgatgatccagtgltattacagtgcttttaa 3095
QY 3061 ctgatactgct 3072
|||||
Db 3096 ctgatactgct 3107

RESULT 2
AAH98254
ID AAH98254 standard; cDNA; 3545 BP.
AC AAH98254;
XX
XX
XX 12-OCT-2001 (first entry)
XX
XX
XX Murine EST-derived coding sequence SEQ ID NO: 111.
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX diagnostics; forensic test; gene mapping; genetic disorder;
XX biodiversity; gene therapy; nutrition; ss.
XX
XX Mus musculus.
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001: 2001MO-US02687.
XX
XX PF
XX 25-JAN-2000: 2000US-0491404.
XX PR 17-JUL-2000: 2000US-0617746.
XX PR 03-AUG-2000: 2000US-0631451.
XX PR 15-SEP-2000: 2000US-0663870.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX PA
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
XX DR WPI: 2001-476164/51.
XX DR P-PSDB; AAM23595.
XX
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX PT antibodies and research use -
XX PT
XX
XX PS Claim 1: Page 250-251; 1275pp; English.

CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
CC
XX
XX SO Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other;

Query Match 95.0%; Score 2919; DB 22; Length 3545;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaattcataaaggacatagccgacctatccaagaatgggaatgactgtata 60
|||||
Db

Db 232 atgaattcataaaggacatagccgacctatccaagaatgggaatgactgtata 291
QY 61 aagcaaatcacagatgacatttgtaatgtaatgcttcgaagcaagaatcaatc 120
|||||
Db 292 aagcaaatcacagatgacatttgtaatgtaatgcttcgaagcaagaatcaatc 351
QY 121 atttgctgcagagaagtgagcagagatgctgtcagaaggatcatcaatgatattgaa 180
|||||
Db 352 atttgctgcagagaagtgagcagagatgctgtcagaaggatcatcaatgatattgaa 411
QY 181 aaggttccaaggtccgtgaatcccttcttaaatcccttaagaagtggaactctctta 240
|||||
Db 412 aaggttccaaggtccgtgaatcccttcttaaatcccttaagaagtggaactctctta 471
QY 241 ttccaagacttgatgacaaagctcttttcacagacatccagaagagactggagat 300
|||||
Db 472 ttccaagacttgatgacaaagctcttttcacagacatccagaagagactggagat 551
QY 301 ttgctcagatcttaagaagactgtacacatcccatctttctgaactttatccctt 360
|||||
Db 532 ttgctcagatcttaagaagactgtacacatcccatctttctgaactttatccctt 591
QY 361 ggtgaagatatgacattatlttaacttgaagaagacacttcacagactgtccgttg 420
|||||
Db 592 ggtgaagatatgacattatlttaacttgaagaagacacttcacagactgtccgttg 651
QY 421 aggaagaccacaac 480
|||||
Db 652 aggaagaccacaac 711
QY 481 ctccaagagccctgcacatcatgaaagggaactctgcgaagaagctccactctgtcag 540
|||||
Db 712 ctccaagagccctgcacatcatgaaagggaactctgcgaagaagctccactctgtcag 771
QY 541 cgcattgcacatgctctggtggtctccggaagtgcaaggtctgcgaagtccaatctgtc 600
|||||
Db 772 cgcattgcacatgctctggtggtctccggaagtgcaaggtctgcgaagtccaatctgtc 831
QY 601 ttctccctccgtctcagcagagcccgaggtgagacttttaaacctctgtatcaactc 660
|||||
Db 832 ttctccctccgtctcagcagagcccgaggtgagacttttaaacctctgtatcaactc 891
QY 661 ctgatatactctgcgaacatcaagaagcagacatcatggtcactgtcgtgaagtcgag 720
|||||
Db 892 ctgatatactctgcgaacatcaagaagcagacatcatggtcactgtcgtgaagtcgag 951
QY 721 caagaggtctcttctctctctgtatgcttaaatgaaatcaagcccaagatgccagaa 780
|||||
Db 952 caagaggtctcttctctctctgtatgcttaaatgaaatcaagcccaagatgccagaa 1011
QY 781 atcgaagccctgataaaggaagaaaccccgcttcaagaacatgltcatgctcacatcacc 840
|||||
Db 1012 atcgaagccctgataaaggaagaaaccccgcttcaagaacatgltcatgctcacatcacc 1071
QY 841 actgaatgctcagcagac 900
|||||
Db 1072 actgaatgctcagcagac 1131
QY 901 acgaagacacagcccgaggtctcaccgagaagtgctgtacagaagagctgtctgaagc 960
|||||
Db 1132 acgaagacacagcccgaggtctcaccgagaagtgctgtacagaagagctgtctgaagc 1191
QY 961 ttgtgctcacaatlcagaatccaggtgttgtagaatlcataagagccctctctt 1020
|||||
Db 1192 ttgtgctcacaatlcagaatccaggtgttgtagaatlcataagagccctctctt 1251
QY 1021 gttgtatcaactgtgcaatccagatggtgtgaagtgtgtccacttcacacacacacac 1080
|||||
Db 1252 gttgtatcaactgtgcaatccagatggtgtgaagtgtgtccacttcacacacacacac 1311
QY 1081 acgcgttccatacctcttatgatactgtgtgatacagaaaaacacacacacacacacacac 1140
|||||
Db 1312 acgcgttccatacctcttatgatactgtgtgatacagaaaaacacacacacacacacacac 1371

OY 1141 gtgagctgaagtgacttcatctcgaaagcctggaaccactgtcggaaccctagactctggaaggt 1200
|||||
Db 1372 gtgagctgaagtgacttcatctcgaaagcctggaaccactgtcggaaccctagactctggaaggt 1431
OY 1201 gtgtctcccaagtgatcttgaaactgcagaaatgtgtccagcgtagaagaatgttc 1260
|||||
Db 1432 gtgtctcccaagtgatcttgaaactgcagaaatgtgtccagcgtagaagaatgttc 1491
OY 1261 ctgctggaacactgggtctctctgttaataatacagctccaaggttccaagccaagtataa 1320
|||||
Db 1492 ctgctggaacactgggtctctctgttaataatacagctccaaggttccaagccaagtataa 1551
OY 1321 tctcttccaagaatctccagaggtacacagcaggaaggaagactcagcaggtttatgaag 1380
|||||
Db 1552 tctcttccaagaatctccagaggtacacagcaggaaggaagactcagcaggtttatgaag 1611
OY 1381 tctcatgagccagaggaagtgacccaagggaatgttacttgcagaaaaatgtttccatt 1440
|||||
Db 1612 tctcatgagccagaggaagtgacccaagggaatgttacttgcagaaaaatgtttccatt 1671
OY 1441 tggagcatataatccaactatagcagcctgcggtacacactgtgtgtcactgtgtgaa 1500
|||||
Db 1672 tggagcatataatccaactatagcagcctgcggtacacactgtgtgtcactgtgtgaa 1731
OY 1501 gccacacagagctgttatgagacccctcgagcagtgatatacaacagcgctgtctcgga 1560
|||||
Db 1732 gccacacagagctgttatgagacccctcgagcagtgatatacaacagcgctgtctcgga 1791
OY 1561 cttccatctgcacaagagcctctctgtgagacaggaatcttgcagaaatgttgaaaaaacc 1620
|||||
Db 1792 cttccatctgcacaagagcctctctgtgagacaggaatcttgcagaaatgttgaaaaaacc 1851
OY 1621 acttgagcaagaatctctgaaagccataaaccatcaatctctgtgagatgtgtgacacat 1680
|||||
Db 1852 acttgagcaagaatctctgaaagccataaaccatcaatctctgtgagatgtgtgacacat 1911
OY 1681 ttatatcaagagtagatataccaatcaagccctggaagccaagaatttgaaccttcttcaa 1740
|||||
Db 1912 ttatatcaagagtagatataccaatcaagccctggaagccaagaatttgaaccttcttcaa 1971
OY 1741 ggtlaaagctatataatacaactcagggaaacatcccgatcttacttacttgaactcttga 1800
|||||
Db 1972 ggtlaaagctatataatacaactcagggaaacatcccgatcttacttacttgaactcttga 2031
OY 1801 catctggccaatctgagcaagtgtctgtgactcattcaaatctgactttatgtgggaagct 1860
|||||
Db 2032 catctggccaatctgagcaagtgtctgtgactcattcaaatctgactttatgtgggaagct 2091
OY 1861 atggtcttctggaagaaagctgacagaagacaggtggaatccacatggaagagccca 1920
|||||
Db 2092 atggtcttctggaagaaagctgacagaagacaggtggaatccacatggaagagccca 2151
OY 1921 gaaacctacatctccagcagagcgtatcttgttcttcaactggaagcaggaatcag 1980
|||||
Db 2152 gaaacctacatctccagcagagcgtatcttgttcttcaactggaagcaggaatcag 2211
OY 1981 actctggaggtacacactccggagatttcagcaagttgaaataagcaagatatcacatctg 2040
|||||
Db 2212 actctggaggtacacactccggagatttcagcaagttgaaataagcaagatatcacatctg 2271
OY 2041 gggaaataatcagactctgcacaagccctcagcgtgcaataaagaagatgtgtgtgtg 2100
|||||
Db 2272 gggaaataatcagactctgcacaagccctcagcgtgcaataaagaagatgtgtgtgtgtg 2331
OY 2101 gctggaagcctcagttgtgtctcagcacctgtlaagaacattatctctcatgtgtgaa 2160
|||||
Db 2332 gctggaagcctcagttgtgtctcagcacctgtlaagaacattatctctcatgtgtgaa 2391
OY 2161 gccagttccctacatagaagaatgagagccatcacactctgttaacaacttgaaacc 2220
|||||
Db 2392 gccagttccctacatagaagaatgagagccatcacactctgttaacaacttgaaacc 2451

OY 2221 ttgagttcatgacctacagaatacaacgctgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2280
|||||
Db 2452 ttgagttcatgacctacagaatacaacgctgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2511
OY 2281 aacttgaagaaccttacaagctcatalgtgaataaagaatgaatgaagaatgtct 2340
|||||
Db 2512 aacttgaagaaccttacaagctcatalgtgaataaagaatgaatgaagaatgtct 2571
OY 2341 ataaactagctggaagccttgaaaaaccctggaagaatgtgtttatttcaatttgaccac 2400
|||||
Db 2572 ataaactagctggaagccttgaaaaaccctggaagaatgtgtttatttcaatttgaccac 2631
OY 2401 ttgtctgacatttgaagaagggaatgtgatacatagatcaagttctctgtccaagtgaacctgt 2460
|||||
Db 2632 ttgtctgacatttgaagaagggaatgtgatacatagatcaagttctctgtccaagtgaacctgt 2691
OY 2461 gaccttgaagaatltcaatltactctctgtctgttcttgcgcaaatgtcagtgaaaaatccta 2520
|||||
Db 2692 gaccttgaagaatltcaatltactctctgtctgttcttgcgcaaatgtcagtgaaaaatccta 2751
OY 2521 gctcagaatcttccacaatttgttccaactgagcatctctgtattatccagaataatccctg 2580
|||||
Db 2752 gctcagaatcttccacaatttgttccaactgagcatctctgtattatccagaataatccctg 2811
OY 2581 gaaaaagatggaatgaagctctctcaatgaactgacatcgacaggaatgaacgtgtctagaagag 2640
|||||
Db 2812 gaaaaagatggaatgaagctctctcaatgaactgacatcgacaggaatgaacgtgtctagaagag 2871
OY 2641 ctcaaccgactgtatgtctgccttgaggctgtgacgtgtgcaagagcagccttgagcagcgtgtg 2700
|||||
Db 2872 ctcaaccgactgtatgtctgccttgaggctgtgacgtgtgcaagagcagccttgagcagcgtgtg 2931
OY 2701 aaacatttgaagaagggtccccaactcgtccaagctgtgggttgaaaaaactggaagctcaca 2760
|||||
Db 2932 aaacatttgaagaagggtccccaactcgtccaagctgtgggttgaaaaaactggaagctcaca 2991
OY 2761 gatatacgaatatagaattttagtgcatttcttgaagaagaccctctgaagaactccag 2820
|||||
Db 2992 gatatacgaatatagaattttagtgcatttcttgaagaagaccctctgaagaactccag 3051
OY 2821 cagttgaatttgcgaggaatcgtgtgacagtgatgagatgtgcttgccttcaatgtgtgtga 2880
|||||
Db 3052 cagttgaatttgcgaggaatcgtgtgacagtgatgagatgtgcttgccttcaatgtgtgtga 3111
OY 2881 ttggagaacttaagaatatagtttttttagactttagacttaagaagaatttactactgt 2940
|||||
Db 3112 ttggagaacttaagaatatagtttttttagactttagacttaagaagaatttactactgt 3171
OY 2941 ccagcatagtcagaagaactttagccaagtgatataccaagttaactttctgcaagaagct 3000
|||||
Db 3172 ccagcatagtcagaagaactttagccaagtgatataccaagttaactttctgcaagaagct 3231
OY 3001 aggcctgtgtggtggaatttgatgatagtatcagtgatatacagtggtctttaa 3060
|||||
Db 3232 aggcctgtgtggtggaatttgatgatagtatcagtgatatacagtggtctttaa 3291
OY 3061 ctgagtaactgtc 3072
|||||
Db 3292 ctgagtaactgtc 3303

RESULT 3
AAS03946
ID AAS03946 standard: DNA; 3615 BP.
XX
XX AAS03946;
XX
XX
DT 12-SEP-2001 (first entry)
XX
XX
DE Human caspase recruitment domain 12 (CARD-12) genomic DNA.
XX
XX Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;

[illegible]

QY	2603	ttcatgaactcgtacgcagatgtaacgtgtcagaaacgtccacgcacatgactgtccct	2662
Db	3074	ttcatgaactcgtacgcagatgtaacgtgtcagaaacgtccacgcacatgactgtccct	3133
QY	2663	ggggctgtgacgtgtaagaagcagccctggaacacgtgttaaacatttggagaaggtccac	2722
Db	3134	ggggctgtgacgtgtaagaagcagccctggaacacgtgttaaacatttggagaaggtccac	3193
QY	2723	aactcgtcaacgcttgggttggaaaaaactggaagactcacagatcacagatgtaattttag	2782
Db	3194	aactcgtcaacgcttgggttggaaaaaactggaagactcacagatcacagatgtaattttag	3253
QY	2783	gtgcatttttttggaaagaacacctctgtaaaaaactccacagatltgaatttggcggaaatc	2842
Db	3254	gtgcatttttttggaaagaacacctctgtaaaaaactccacagatltgaatttggcggaaatc	3313
QY	2843	gtgtgaagaagatgtagtgaagcttgcttcacatgggtgtgatttggaaatcttaagaattag	2902
Db	3314	gtgtgaagaagatgtagtgaagcttgcttcacatgggtgtgatttggaaatcttaagaattag	3373
QY	2903	tgtttttgactttgactaactaaagaattctaccctgcagatccagacttagtcagaactta	2962
Db	3374	tgtttttgactttgactaactaaagaattctaccctgcagatccagacttagtcagaactta	3433
QY	2963	gccaaagtgttatccaaagttaactttctgcagaagaagctagagcttgggtgtgcgaatttg	3022
Db	3434	gccaaagtgttatccaaagttaactttctgcagaagaagctagagcttgggtgtgcgaatttg	3493
QY	3023	atgatcatgatctcagttgatttacaag	3049
Db	3494	atgatcatgatctcagttgatttacaag	3520
RESULT 4			
AAH99581			
ID	AAH99581	standard; cDNA; 2950 BP.	
AC	AAH99581:		
XX			
DT	16-OCT-2001	(first entry)	
XX			
DE	Human protein encoding cDNA sequence SEQ ID NO:416.		
XX			
KM	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;		
KM	antiflammatory; antirheumatic; antirheumatic; antirheumatic; antirheumatic;		
KM	antibacterial; endocrine; cardiac; central nervous system; vitruclide;		
KM	anti-HIV; fungicide; antimetastatic; cardiovascular; antianemic; anemia;		
KM	antiaagregant; haemostatic; vulnery; antileuk; osteopathic; eczema;		
KM	dermatological; antiallergic; antiaesthetic; antidiabetic; cyclostatic;		
KM	neuroprotective; antidepressant; nocotropic; antiparkinsonian; infection		
KM	immunostimulant; gene therapy; antisense therapy; vaccine; inflammation		
KM	antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;		
KM	cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;		
KM	genetic disease; haematopoietic disorder; platelet disorder; asthma;		
KM	thrombocytopenia; osteoporosis; severe combined immunodeficiency;		
KM	allergic rhinitis; diabetes; multiple sclerosis; depression;		
KM	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;		
KM	neurological disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153455-A2.		
XX			
PD	26-JUL-2001.		
XX			
PF	22-DEC-2000; 2000WO-US35017.		
XX			
PR	23-DEC-1999; 99US-0471275.		
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
XX			
PA	(HYSE-) HYSEQ INC.		

(HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI: 2001-457603/49.
DR P-PSDB; AAM25640.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 1, Page 511-512; 1217pp; English.
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antineurotic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; viricide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antineoplastic; antiagregant; haemostatic; vulnerary;
CC anticancer; osteopathic; dermatological; antiallergic; antisthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other;
QY
Query Match 58.0%; Score 1781; DB 22; Length 2950;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1881; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1190 cctctgagagtggtctctccacaagttgatttcgaactcagagatgttccagcgtga 1249
Db 45 cctctgagagtggtctctccacaagttgatttcgaactcagagatgttccagcgtga 104
QY 1250 atgagagatgtcctcgtcgaacactggcctcctgtlaaatalacagctccaaggttcaagc 1309
Db 105 atgagagatgtcctcgtcgaacactggcctcctcgtlaaatalacagctccaaggttcaagc 164
QY 1310 caaagataaatctcttcaaaagtcattccagaagatatacaagaagcagaagactcagca 1369
Db 165 caaagataaatctcttcaaaagtcattccagaagatatacaagaagcagaagactcagca 224
QY 1370 gttatctgacgtctcatgagccagagagtgagcacaagggaaatgtgttacttcaagaaa 1429
Db 225 gttatctgacgtctcatgagccagagagtgagcacaagggaaatgtgttacttcaagaaa 284
QY 1430 tggttccattcctggaatatacaactatagcagctcgtccggttacacctgttgggt 1489
Db 285 tggttccattcctggaatatacaactatagcagctcgtccggttacacctgttgggt 344
QY 1490 catctgtggaagcaccagagcgtgttatgaagcactcagcagatgtatcaacaagcgtc 1549
Db 345 catctgtggaagcaccagagcgtgttatgaagcactcagcagatgtatcaacaagcgtc 404
QY 1550 gaccttcgaccttcacatcgccaaagagcctctctgagagcaggaattcttcaaatg 1609
Db 405 gaccttcgaccttcacatcgccaaagagcctctctgagagcaggaattcttcaaatg 464
QY 1610 tgaanaaacaccatgagcaagaatattctgaaagccataaacaatcaatctcttggtagag 1669
Db 465 tgaanaaacaccatgagcaagaatattctgaaagccataaacaatcaatctcttggtagag 524
QY 1670 gtggcatcattatatacaagagatatacatcaaatcaagcctcgtgagcagaatttgaag 1729

Db 525 gtggcatcattatatacaagagatatacatcaaatcaagcctcgtgagcagaatttgaag 584
QY 1730 cttctctcaaggttaaaggttatatacaactcaaggaacatcccgatattatttg 1789
Db 585 cttctctcaaggttaaaggttatatacaactcaaggaacatcccgatattatttg 644
QY 1790 actctctgacacattgcccacattgtgcaagtgctcgtgacatcaatactgacttt 1849
Db 645 actctctgacacattgcccacattgtgcaagtgctcgtgacatcaatactgacttt 704
QY 1850 atggggagatagcttctcatgtggaaaagctgcagaagcacaagttgtaattccaatgg 1909
Db 705 atggggagatagcttctcatgtggaaaagctgcagaagcacaagttgtaattccaatgg 764
QY 1910 aaagagcccccagaacactaatattccagcagggcttatatttggcttccaactggaagc 1969
Db 765 aaagagcccccagaacactaatattccagcagggcttatatttggcttccaactggaagc 824
QY 1970 aggaattcaggaactcgtgaggtcacaactccggatctcagaagtgtgaataagaagata 2029
Db 825 aggaattcaggaactcgtgaggtcacaactccggatctcagaagtgtgaataagaagata 884
QY 2030 tccacatctggggaaaatatattagctctgcccagaagcctcagctgcacaataaagaat 2089
Db 885 tccacatctggggaaaatatattagctctgcccagaagcctcagctgcacaataaagaat 944
QY 2090 gtctctgtgtgtgtggaagcctcagtttggcttcctcagcacttgaagacattattctc 2149
Db 945 gtctctgtgtgtgtggaagcctcagtttggcttcctcagcacttgaagacattattctc 1004
QY 2150 tcatgtgtggaagcagctccctccacatagaagaatgagagcacaatcacatctgtlaaca 2209
Db 1005 tcatgtgtggaagcagctccctccacatagaagaatgagagcacaatcacatctgtlaaca 1064
QY 2210 acctgtaaaacctgtgattatcagactcaagaatataagcgtgcgggtgtgtctgacg 2269
Db 1065 acctgtaaaacctgtgattatcagactcaagaatataagcgtgcgggtgtgtctgacg 1124
QY 2270 acagctgtgtlaacttgaagaaccttacaagaactcaatgaatgaatgaatgaatgaatga 2329
Db 1125 acagctgtgtlaacttgaagaaccttacaagaactcaatgaatgaatgaatgaatgaatga 1184
QY 2330 aagaagaatgctataaactagctgtaagggcctgtaaaaacctgtaagaagaatgtgttattcc 2389
Db 1185 aagaagaatgctataaactagctgtaagggcctgtaaaaacctgtaagaagaatgtgttattcc 1244
QY 2390 attgagccacctgtgtgacattgagagaggaatgtgattatcaatgaatcctctgtcaca 2449
Db 1245 attgagccacctgtgtgacattgagagaggaatgtgattatcaatgaatcctctgtcaca 1304
QY 2450 gtgaacctgtgaccttgaagaatcaatcaattagctcctcgtctgtctgcgaatgacg 2509
Db 1305 gtgaacctgtgaccttgaagaatcaatcaattagctcctcgtctgtctgcgaatgacg 1364
QY 2510 tgaanaactagctcagaatcttcaaatatttggatcaactgaagcattctgattatcag 2569
Db 1365 tgaanaactagctcagaatcttcaaatatttggatcaactgaagcattctgattatcag 1424
QY 2570 aaaaattacacctggaanaagatggaatgaagctcttcatgaactcagacagaatgaag 2629
Db 1425 aaaaattacacctggaanaagatggaatgaagctcttcatgaactcagacagaatgaag 1484
QY 2630 tgcataaagcctcacagcactgatactgcccctggggcctgtgacgtgcaagcagcctga 2689
Db 1485 tgcataaagcctcacagcactgatactgcccctggggcctgtgacgtgcaagcagcctga 1544
QY 2690 gcagcctgttgaacaacttggagaggtcccaacaactcgtcaagcctgtgtgtaaaaact 2749
Db 1545 gcagcctgttgaacaacttggagaggtcccaacaactcgtcaagcctgtgtgtaaaaact 1604
QY 2750 ggaagctcaagatatacagagattagaatttgaagtgcaatlltttggaaagaacctctga 2809

DB 1605 ggagactcacagatagagattagaatttagtgatcttlttggaaagaccctctga 1664
 QY 2810 aaacttcagagtgtaatttggcgggaatcgtgtgacgtatgatgtcttgcct 2869
 DB 1665 aaacttcagagtgtaatttggcgggaatcgtgtgacgtatgatgtcttgcct 1724
 QY 2870 tcattggttatttgagatccttaagcaattatgttlttggacttgaacaaagaat 2929
 DB 1725 tcattggttatttgagatccttaagcaattatgttlttggacttgaacaaagaat 1784
 QY 2930 tctactatccagcagcattagtcagaanaacttagccaagtgtatccaaagttaacttcc 2989
 DB 1785 tctactatccagcagcattagtcagaanaacttagccaagtgtatccaaagttaacttcc 1844
 QY 2990 tgcagaagcctgagctgtgttgggtgacatttgatgatgtatccatgttataacag 3049
 DB 1845 tgcagaagcctgagctgtgttgggtgacatttgatgatgtatccatgttataacag 1904
 QY 3050 gtgctttaactagtaactgct 3072
 DB 1905 gtgctttaactagtaactgct 1927

RESULT 5
 AAH34171/C
 ID AAH34171 standard; cDNA; 2735 BP.
 XX
 AC AAH34171;
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1253.
 XX
 KM Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KM colorectal carcinoma; ss.
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 DR WPT. 2001-235357/24.
 DR P-PSDB: AAG74766.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1; Page 3017; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytosolic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAH77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SO Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other;

Query Match 15.9%; Score 489; DB 22; Length 2735;
 Best Local Similarity 99.8%; Pred. No. 1.8e-232;
 Matches 539; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2533 cacaatttggccaactggagcatcttgatttaccagaaataactctggaagaatgga 2592
 DB 2727 CACAATTTCGTCAACTGAGCATCTTGATTTCAGAAAATTACCTCGAAAAAGATGGA 2668
 QY 2593 aatgaagctcttcatactgatactcgacagatgaagctgtagaagactccaccgactg 2652
 DB 2667 AATGAAGCTCTTCATGATGATCGACAGATGAACGTGCTAGAACACTCACCGACTG 2608
 QY 2653 atgtgcacctggggcctgtgacgtgcaaggcaagccttgagcagcctgttgaacaattgag 2712
 DB 2607 ATGCTGCCCTGGGGCTGTGACGTGCAAGCAGCGCTGAGCAGCTTGTGAACATTGGAG 2548
 QY 2713 gaggctcccaactcgttcaagcttgggttgaanaacttgagactcacagatagatt 2772
 DB 2547 GAGGTCCTCACAACTGCTCAAGCTTGGGTTGAAAACCTGAGACTCACAGATACGAGATT 2488
 QY 2773 agaatlttagtgatcttlttggaaagaaccccttgaanaacttcagcagtgtaattg 2832
 DB 2487 AGAATTTAGTGTCATTTTGGAAAGAACCCCTGTAATAAATCTCCAGACTTGAATTGG 2428
 QY 2833 gcgggaatcgtgtgagcagtgatgtgctgtccctcaagggtgttattggaactct 2892
 DB 2427 GCGGGAATCGTGCTGAGCAGTGTGATGATGCTGCTTCATGAGGTATTTGAGAACTT 2368
 QY 2893 aagcaattagtgatcttlttggacttgaacttactaaagaatttctactgataccagatagtc 2952
 DB 2367 AAGCAATTAGTGTTTTGTGACTTACTTAAGAAATTTCTACTGATCCAGCATTAAGTC 2308
 QY 2953 agaaacttagccaagtgatcccaagtttaacttcttcgaagaagctagagctgttgg 3012
 DB 2307 AGAAACTTAGCCAAAGTGTATCCAGTTACTTTCTGCAAGAACTAGGCTTGTGGG 2248
 QY 3013 tggcaatttgatgatgatcctcagtgatatacacaggtgtctttaaactagtaactgct 3072
 DB 2247 TGGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2188

RESULT 6
 AA114389
 ID AA114389 standard; DNA; 421 BP.
 XX
 AC AA114389;
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #4322 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 4322; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.9%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. NO. 7.6e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2781 agtgcatttttggaaagaccctcgtgaaacctccagcaattgagtcgggaaa 2840
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 180 aggtgcatttttggaaagaccctcgtgaaacctccagcaattgagtcgggaaa 239
QY 2841 tcgtgagcagtgatgagtgcttcgctcctcagtggtgtatcttgagaatcttaagcaat 2900
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 240 tcgtgagcagtgatgagtgcttcgctcctcagtggtgtatcttgagaatcttaagcaat 299
QY 2901 agtgcatttttggacttaagtaagaattctaccgtatccagcatltagtcagaaaact 2960
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 300 agtgcatttttggacttaagtaagaattctaccgtatccagcatltagtcagaaaact 359
QY 2961 taagcaagtgatataccaagttaactttctgcaagaagcaggctgttggtggcaat 3020
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 360 taagcaagtgatataccaagttaactttctgcaagaagcaggctgttggtggcaat 419
QY 3021 tg 3022
DB ||
DB 420 tg 421

RESULT 7
AAI35764
ID AAI35764 standard; DNA; 421 BP.
XX
AC AAI35764;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #4450 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200157272-A2.
XX
XX

PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 4450; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.9%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. NO. 7.6e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2781 agtgcatttttggaaagaccctcgtgaaacctccagcaattgagtcgggaaa 2840
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 180 aggtgcatttttggaaagaccctcgtgaaacctccagcaattgagtcgggaaa 239
QY 2841 tcgtgagcagtgatgagtgcttcgctcctcagtggtgtatcttgagaatcttaagcaat 2900
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 240 tcgtgagcagtgatgagtgcttcgctcctcagtggtgtatcttgagaatcttaagcaat 299
QY 2901 agtgcatttttggacttaagtaagaattctaccgtatccagcatltagtcagaaaact 2960
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 300 agtgcatttttggacttaagtaagaattctaccgtatccagcatltagtcagaaaact 359
QY 2961 taagcaagtgatataccaagttaactttctgcaagaagcaggctgttggtggcaat 3020
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 360 taagcaagtgatataccaagttaactttctgcaagaagcaggctgttggtggcaat 419
QY 3021 tg 3022
DB ||
DB 420 tg 421

RESULT 8
AAI04213
ID AAI04213 standard; DNA; 421 BP.
XX
AC AAI04213;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4204 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
XX WO200157270-A2.
XX
XX

```

XX 09-AUG-2001.
PD 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 25; SEQ ID No 4204; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
XX
Query Match          7.9%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. No. 7.6e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2781 aggtgcatttttggaaagaccctctgaaacttcagcagcttgaattggcgga 2840
   |||||||
DB 180 aggtgcatttttggaaagaccctctgaaacttcagcagcttgaattggcgga 239
QY 2841 tctgtgaacagtgatgagctgtcgtcccttcagtgagtgatctgaagaacttcgaagaatt 2900
   |||||||
DB 240 tctgtgaacagtgatgagctgtcgtcccttcagtgagtgatctgaagaacttcgaagaatt 299
QY 2901 agtgttttttgacttaagacttaagactttctactgtatccagcaacttaagcaact 2960
   |||||||
DB 300 agtgttttttgacttaagacttaagactttctactgtatccagcaacttaagcaact 359
QY 2961 tagcgaagtgatcccaagcttaactttctgcaagaagctaggtctgtgggtggcaatt 3020
   |||||||
DB 360 tagcgaagtgatcccaagcttaactttctgcaagaagctaggtctgtgggtggcaatt 419
QY 3021 tg 3022
   ||
DB 420 tg 421

```

```

XX DE Probe #13523 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX XX WO200157278-A2.
XX PN
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 13523; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
XX
Query Match          7.2%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2830 ttcgctgggaactcgtgtgacagtgatgagctgtcctcctaagtggtattgagaat 2889
   |||||||
DB 1 ttcgctgggaactcgtgtgacagtgatgagctgtcctcctaagtggtattgagaat 60
QY 2890 cttaagcaattggttttttgacttaagacttaagaaactttctactgataccagcata 2949
   |||||||
DB 61 cttaagcaattggttttttgacttaagacttaagaaactttctactgataccagcata 120
QY 2950 gtccagaaacttagccaagtgtatccaaagcttaactttctgcaagaagctaggtgtt 3009
   |||||||
DB 121 gtccagaaacttagccaagtgtatccaaagcttaactttctgcaagaagctaggtgtt 180
QY 3010 gggctggcaatttgatgatgatcctcagtgattatacag 3049
   |||||||
DB 181 gggctggcaatttgatgatgatcctcagtgattatacag 220

```



```
XX 17-OCT-2001 (first entry)
XX
XX Probe #17590 used to measure gene expression in human placenta sample.
DE
XX
XX Probe: microarray: human; placenta; antenatal diagnosis:
KM genetic disorder; ss.
XX
XX Homo sapiens:
OS
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 17590; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other:
SQ
Query Match 7.2%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2830 ttggcggaatggtgtgacagatgagatgagtgctgcttcattggtgatttgaat 2889
Db 1 ttggcggaatggtgtgacagatgagatgagtgctgcttcattggtgatttgaat 60
QY 2890 cttaagcaattagtgcttttctgacttagtactaaagaatttctaccgtaccagcat 2949
Db 61 cttaagcaattagtgcttttctgacttagtactaaagaatttctaccgtaccagcat 120
QY 2950 gtcaagaacttagccaagtgtatccaaagttaactttctgcaagaagctagctgtt 3009
Db 121 gtcaagaacttagccaagtgtatccaaagttaactttctgcaagaagctagctgtt 180
QY 3010 ggggtgcaatttgatgatgatgtctcagtgattatcacag 3049
Db 181 ggggtgcaatttgatgatgatgtctcagtgattatcacag 220
RESULT 11
AA109206
ID AA109206 standard; DNA: 220 BP.
XX
XX AA109206;
XX
XX 09-OCT-2001 (first entry)
XX
```

```
DE Probe #9197 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 25; SEQ ID No 9197; 322bp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridizes at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other:
SQ
Query Match 7.2%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2830 ttggcggaatggtgtgacagatgagatgagtgctgcttcattggtgatttgaat 2889
Db 1 ttggcggaatggtgtgacagatgagatgagtgctgcttcattggtgatttgaat 60
QY 2890 cttaagcaattagtgcttttctgacttagtactaaagaatttctaccgtaccagcat 2949
Db 61 cttaagcaattagtgcttttctgacttagtactaaagaatttctaccgtaccagcat 120
QY 2950 gtcaagaacttagccaagtgtatccaaagttaactttctgcaagaagctagctgtt 3009
Db 121 gtcaagaacttagccaagtgtatccaaagttaactttctgcaagaagctagctgtt 180
QY 3010 ggggtgcaatttgatgatgatgtctcagtgattatcacag 3049
Db 181 ggggtgcaatttgatgatgatgtctcagtgattatcacag 220
RESULT 12
AAH11452
ID AAH11452 standard; cDNA: 549 BP.
XX
```


XX AAH1452;
 AC 26-JUN-2001 (first entry)
 DT Human CDNA clone (3'-primer) SEQ ID NO:8287.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW Homo sapiens.
 XX EP1074617-A2.
 PN 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 PE 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length CDNA defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length CDNA -
 PS Claim 3; SEQ ID 8287; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length CDNA defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length CDNA. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length CDNA. The primers allow obtaining of the full-length
 CC CDNA easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 549 BP; 162 A; 120 C; 119 G; 139 T; 9 other;

Query Match 0.6%; Score 19; DB 22; Length 549;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 caaagaatgggaatgactg 55
 |||||
 DB 214 caaagaatgggaatgactg 232

RESULT 13

AAH30628/C
 ID AAX30628 standard; DNA; 579 BP.
 XX AAX30628;
 AC 08-JUN-1999 (first entry)
 DT H. pylori cell envelope protein ORF hp7e10433_5345837_c2_8.
 DE H. pylori cell envelope protein ORF; cell envelope protein;
 KW Vaccines; probe; diagnostic; ORF; secreted protein; cytoplasmic protein; cellular protein; ds.
 XX Helicobacter pylori.
 OS WO9824475-A1.
 PN 11-JUN-1998.
 PD 05-DEC-1997; 97WO-US22104.
 PE 14-JUL-1997; 97US-0891928.
 PR 05-DEC-1996; 96US-0753625.
 PR 25-MAR-1997; 97US-0823745.
 XX (ASTR) ASTRA AB.
 PA Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 PI WPI: 1998-333051/29.
 DR P-PSDB; AAY11099.
 XX New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PT H. pylori and other Helicobacter species
 PS Claims 3, 4; Page 169; 339pp; English.
 XX Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 XX Sequence 579 BP; 173 A; 130 C; 98 G; 178 T; 0 other;

Query Match 0.6%; Score 19; DB 19; Length 579;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 acattatttttaactgaa 392
 |||||
 DB 356 ACATTATTTTAACTTGAA 338

RESULT 14
 ID AAX14445/C
 AC AAX14445 standard; DNA; 649 BP.
 XX AAX14445;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPO 875 gene.
 XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease; ss.

```

XX OS Helicobacter pylori.
XX
XX Key Location/Qualifiers
FH CDS 19..624
FT /*tag= a
FT
XX
XX WO9843478-A1.
XX
XX PD 08-OCT-1998.
XX
XX PE 01-APR-1998: 98WO-US06371.
XX
XX PR 29-JUL-1997: 97US-0902615.
XX PR 01-APR-1997: 97US-0833457.
XX PR 24-JUN-1997: 97US-0881227.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
XX DR WPI: 1998-542293/46.
XX DR P-PSDB; AAM98726.
XX
XX PT New isolated Helicobacter polynucleotides - used to develop products
XX PT for the diagnosis, prevention and treatment of Helicobacter
XX PT infections and gastrointestinal diseases
XX
XX PS Claim 1: Page 1538: 2054pp; English.
XX
XX CC This sequence represents a polynucleotide of the invention. It was
XX CC isolated from Helicobacter pylori and encodes a H.pylori GHPo protein.
XX CC The polypeptides can be used for preventing or treating Helicobacter
XX CC infections, and gastroduodenal diseases associated with these
XX CC infections, including acute, chronic, and atrophic gastritis, and peptic
XX CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
XX CC for the production of antibodies. The products can also be used for
XX CC detection and diagnosis.
XX
XX SQ Sequence 649 BP: 191 A; 139 C; 118 G; 201 T; 0 other;

Query Match 0.6%; Score 19; DB 19; Length 649;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 acattatttacttgaa 392
   ||||||||||||||||
Db 401 ACAATATTTTAACTTGA 383

RESULT 15
AAFI3098
ID AAFI3098 standard; cDNA; 1908 BP.
XX
XX AC AAFI3098;
XX
XX DT 13-MAR-2001 (first entry)
XX
XX DE Aspergillus oryzae EST SEQ ID NO:5621.
XX
XX KW Multiple gene expression; filamentous fungal cell; EST;
XX KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX KW culture condition; environmental stress; spore morphogenesis;
XX KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX OS Aspergillus oryzae.
XX
XX PN WO200056762-A2.
XX
XX PD 28-SEP-2000.

```

```

XX XX 22-MAR-2000: 2000WO-US07781.
XX PF
XX PR 22-MAR-1999: 99US-0273623.
XX
XX PA (NOVO) NOVO NORDISK BIOTECH INC.
XX PA (NOVO) NOVO NORDISK AS.
XX
XX PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
XX PI WPI: 2000-594572/56.
XX
XX DR
XX
XX PT Monitoring differential expression of genes in filamentous fungal cells
XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PT substrate of expressed sequence tags -
XX
XX PS Claim 88: Page 2332: 3161pp; English.
XX
XX CC The present invention describes a method for monitoring differential
XX CC expression of genes in a first filamentous fungal (FF) cell relative to
XX CC expression of the same genes in one or more second filamentous fungal
XX CC cells. The method uses fluorescence-labeled nucleic acids isolated from
XX CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX CC are used in the methods for monitoring differential expression of genes
XX CC in a first filamentous fungal (FF) cell relative to expression of the
XX CC same genes in one or more second filamentous fungal cells. Monitoring
XX CC the global expression of genes from FF cells allows the production
XX CC potential of the microorganisms to be improved. New genes may be
XX CC discovered, possible functions of unknown open reading frames can be
XX CC identified and gene copy number variation and stability can be
XX CC monitored. The expression of genes can be used to study how FF cells
XX CC adapt to changes in culture conditions, environmental stress, spore
XX CC morphogenesis, recombination, metabolic or catabolic pathway
XX CC engineering. Using ESTs provides several advantages over genomic or
XX CC random cDNA clones including elimination of redundancy as one spot on an
XX CC array equals one gene or open reading frame, and organisation of the
XX CC microarrays based on function of the gene products to facilitate
XX CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
XX CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
XX CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
XX CC all specifically claimed in the present invention.
XX
XX SQ Sequence 1908 BP: 441 A; 497 C; 466 G; 496 T; 8 other;

Query Match 0.6%; Score 19; DB 21; Length 1908;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 ggcagtcacactcgtctgc 538
   ||||||||||||||||
Db 883 ggcagtcacactcgtctgc 901

RESULT 16
AAHI6202/C
ID AAHI6202 standard; cDNA; 2825 BP.
XX
XX AC AAHI6202;
XX
XX DT 26-JUN-2001 (first entry)
XX
XX DE Human cDNA sequence SEQ ID NO:15005.
XX
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX KW Homo sapiens.
XX
XX OS
XX PN EP1074617-A2.
XX
XX PD 07-FEB-2001.
XX

```

PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0185767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 15005; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2825 BP; 774 A; 585 C; 582 G; 884 T; 0 other.

Query Match 0.6%; Score 19; DB 22; Length 2825;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 caaagaatgggaatgactg 55
 |||||||||||||||||||||
 DB 2612 CAAAGAATGGGAATGACTG 2594

RESULT 17
 AAC14986
 ID AAC14986 standard; cDNA: 116 BP.
 XX
 AC AAC14986;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 19061.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PA EPI033401-A2.
 XX

XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 19061; 71bp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dr primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 116 BP; 48 A; 18 C; 25 G; 25 T; 0 other.

Query Match 0.6%; Score 18; DB 21; Length 116;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 agcaaatcacagatgacc 79
 |||||||||||||||||||
 DB 47 agcaaatcacagatgacc 64

RESULT 18
 AAA31261/C
 ID AAA31261 standard; DNA: 274 BP.
 XX
 AC AAA31261;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #222.
 XX
 KW Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO9967421-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 25-JUN-1999; 99WO-NZ00092.
 XX
 PR 25-JUN-1998; 98US-0105307.
 XX
 PA (GENE-) GENESTS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX

PI Havukkala JJ, Bloksberg LN, Glenn M;
 XX WPI: 2000-116958/10.
 DR
 XX
 PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 XX
 XX Claim 1: Page 140-141; 392pp; English.
 CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 XX
 SQ Sequence 274 BP; 59 A; 79 C; 75 G; 60 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 274;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3021 tgaatgatgatcagcag 3038
 DB 101 TGATGATGATGATCTCAG 84

RESULT 19
 AAT24250/C
 ID AAT24250 standard; cDNA to mRNA; 322 BP.
 XX
 AC AAT24250;
 XX
 DT 16-OCT-1996 (first entry)
 XX
 DE Human gene signature HUMGS06268.
 XX
 OS Homo sapiens.
 XX
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PE 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI: 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1: Page 1566; 2245pp; Japanese.
 XX
 CC A singe-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX
 SQ Sequence 322 BP; 84 A; 53 C; 55 G; 119 T; 11 other;

Query Match 0.6%; Score 18; DB 16; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgaattcataaagac 18
 DB 51 ATGAATTCATTAAGAC 34

RESULT 20
 AAA31416/C
 ID AAA31416 standard; DNA; 355 BP.
 XX
 AC AAA31416;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #377.
 XX
 OS Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO9967421-A1.
 XX
 PD 29-DEC-1999.
 XX
 PE 25-JUN-1999; 99WO-NZ00092.
 XX
 PR 25-JUN-1998; 98US-0105307.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Havukkala JJ, Bloksberg LN, Glenn M;
 XX
 DR WPI: 2000-116958/10.
 XX
 PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 XX
 PS Claim 1: Page 188; 392pp; English.
 XX
 CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design

CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 355 BP; 69 A; 98 C; 119 G; 68 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3021 tgatgatgatgcacg 3038
|||||
DB 101 TGATGATGATGATCAG 84

RESULT 21

AAA31325/C
ID AAA31325 standard; DNA; 377 BP.

AC AAA31325;

DT 05-JUL-2000 (first entry)

DE Plant microsatellite marker #286.

XX Plant microsatellite sequence; core repeat sequence; detection; probe;

KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.

XX Eucalyptus grandis.

XX WO9967421-A1.

XX 29-DEC-1999.

XX 25-JUN-1999; 99WO-N200092.

XX 25-JUN-1998; 98US-0105307.

XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Havukkala IJ, Bloksberg LN, Glenn M;

XX WPI: 2000-116958/10.

XX New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -

PS Claim 1; Page 161; 392pp; English.

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.

XX Sequence 377 BP; 72 A; 102 C; 130 G; 73 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 377;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3021 tgatgatgatgcacg 3038
|||||

DB 88 TGATGATGATGATCAG 71

RESULT 22

AAA31366/C
ID AAA31366 standard; DNA; 435 BP.

AC AAA31366;

DT 05-JUL-2000 (first entry)

DE Plant microsatellite marker #327.

XX Plant microsatellite sequence; core repeat sequence; detection; probe;

KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.

XX Eucalyptus grandis.

XX WO9967421-A1.

XX 29-DEC-1999.

XX 25-JUN-1999; 99WO-N200092.

XX 25-JUN-1998; 98US-0105307.

XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Havukkala IJ, Bloksberg LN, Glenn M;

XX WPI: 2000-116958/10.

XX New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -

PS Claim 1; Page 173; 392pp; English.

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.

XX Sequence 435 BP; 85 A; 125 C; 146 G; 79 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3021 tgatgatgatgcacg 3038
|||||
DB 101 TGATGATGATGATCAG 84

RESULT 23

AAC09333/C
ID AAC09333 standard; cDNA; 446 BP.

AC AAC09333;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 13408.

```

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX EPI033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GENEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1, SEQ ID 13408; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC cDNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or poly+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 446 BP; 134 A; 76 C; 70 G; 162 T; 4 other;

Query Match          0.6%; Score 18; DB 21; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaatttcataagac 18
   |||
DB 106 ATGAATTTCATAAGGAC 89

RESULT 24
AII14905/C
ID AII14905 standard; DNA; 454 BP.
AC AII14905;
XX
XX 12-OCT-2001 (first entry)
XX
DE Probe #4838 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
XX WO200157278-A2.
XX
XX - 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.

```

```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 4838; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match          0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 ggggaatctgcaaaagc 522
   |||
DB 29 GGCGAATCTGCAAAAGC 12

RESULT 25
AII16257/C
ID AII16257 standard; DNA; 454 BP.
AC AII16257;
XX
XX 17-OCT-2001 (first entry)
XX
DE Probe #4943 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX

```

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID NO 4943; 654pp; English.
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;
Query Match 0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 505 ggggaatctggcaagc 522
IIIIIIIIIIIIIIIIIIII
DB 29 GGGGAATCTGCCAAGGC 12
RESULT 26
AAI04685/C
ID AAI04685 standard; DNA; 454 BP.
XX AC AAI04685;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #4676 used to measure gene expression in human breast sample.
XX KW Probe: human; breast disease; breast cancer; development disorder; ss;
XX KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN W0200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast -
XX PS Claim 25; SEQ ID NO 4676; 322pp; English.
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,

CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;
Query Match 0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 505 ggggaatctggcaagc 522
IIIIIIIIIIIIIIIIIIII
DB 29 GGGGAATCTGCCAAGGC 12
RESULT 27
AAC01809
ID AAC01809 standard; CDNA; 466 BP.
XX AC AAC01809;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 1807.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR P-PSDB; AAG01803.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 1807; 71pp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. An ORF has been identified within the
XX CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX CC derived from 30 different tissues. EST sequences usually correspond
XX CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX CC well suited for isolating cDNA sequences derived from the 5' ends of
XX CC mRNAs and even in those cases where longer cDNA sequences have been
XX CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX CC mRNAs with intact 5' ends and can therefore be used to obtain full length
XX CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX CC gene therapy and chromosome mapping procedures. They are used to obtain
XX CC upstream regulatory sequences and to design expression and secretion
XX CC vectors.
XX SQ Sequence 466 BP; 76 A; 106 C; 132 G; 148 T; 4 other;

Query Match 0.6%; Score 18; DB 21; Length 466;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 731 tttctctcttgatgct 748
 ||||||||||||||||
 DB 220 tttctctcttgatgct 237

RESULT 28

AAA31287/c
 ID AAA31287 standard; DNA; 468 BP.

XX AAA31287;
 AC
 XX

DT 05-JUL-2000 (first entry)
 XX

DE Plant microsatellite marker #248.
 XX

XX Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.

OS Eucalyptus grandis.
 XX

PN WO967421-A1.
 XX

PD 29-DEC-1999.
 XX

PF 25-JUN-1999; 99WO-NZ00092.
 XX

PR 25-JUN-1998; 98US-0105307.
 XX

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PA

PI Havukkala IJ, Blosberg LN, Glenn M;
 XX

DR WPI: 2000-116958/10.
 XX

PT New plant microsatellite markers and associated flanking species for
 XX the detection of polymorphic genetic markers -
 PS

Claim 1: Page 149; 392pp; English.
 XX

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 CC

XX Sequence 468 BP; 95 A; 129 C; 156 G; 87 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 468;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3021 tgatgatgatgatcag 3038
 ||||||||||||||||

DB 101 TGATGATGATGATCTCAG 84

RESULT 29

AA111787
 ID AA111787 standard; DNA; 472 BP.

XX AA111787;
 AC
 XX

DT 12-OCT-2001 (first entry)
 XX

DE Probe #1720 for gene expression analysis in human cervical cell sample.
 XX

KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.

OS Homo sapiens.
 XX

PN WO200157278-A2.
 XX

PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00670.
 XX

PR 04-FEB-2000; 2000US-0180312.
 XX

PR 26-MAY-2000; 2000US-0207456.
 XX

PR 30-JUN-2000; 2000US-0608408.
 XX

PR 03-AUG-2000; 2000US-0632266.
 XX

PR 21-SEP-2000; 2000US-0234687.
 XX

PR 27-SEP-2000; 2000US-0236359.
 XX

PR 04-OCT-2000; 2000GB-0024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX

DR WPI: 2001-488901/53.
 XX

PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -
 PS

Claim 25; SEQ ID No 1720; 487pp; English.
 XX

CC The present invention relates to human single exon nucleic acid probes
 CC (SENP). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC

XX Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 131 agaagtgagcagcagtg 148
 ||||||||||||||||

DB 143 agaagtgagcagcagtg 160

RESULT 30

AA133101
 ID AA133101 standard; DNA; 472 BP.

XX AA133101;
 AC
 XX

DT 17-OCT-2001 (first entry)
 XX

DE Probe #1787 used to measure gene expression in human placenta sample.
 XX

KW Probe: microarray; human; placenta; antenatal diagnosis;
 XX

KW genetic disorder; ss.
 XX Homo sapiens.
 OS WO200157272-A2.
 XX
 XX PD 09-AUG-2001.
 XX
 XX PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-488897/53.
 DR
 XX
 XX PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 1787; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 SQ Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 131 agaaggtgagcagcagtg 148
 ||||||||||||||||
 DB 143 agaaggtgagcagcagtg 160

RESULT 31
 AA101718
 ID AA101718 standard; DNA: 472 BP.
 XX
 AC AA101718;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Probe #1709 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 XX WO200157270-A2.
 XX
 XX PD 09-AUG-2001.
 XX
 XX PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-476286/51.
 XX
 XX DR 09-AUG-2001.
 XX
 XX PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 1709; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published-pct-sequences.
 XX
 SQ Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 131 agaaggtgagcagcagtg 148
 ||||||||||||||||
 DB 143 agaaggtgagcagcagtg 160

RESULT 32
 AA005868/C
 ID AA005868 standard; DNA: 594 BP.
 XX
 AC AA005868;
 XX
 DT 07-JAN-1991 (first entry)
 XX
 DE Sequence encoding mammalian growth hormone receptor binding protein.
 XX
 KW Ovine placental lactogen; ds;
 KW
 XX Ovis ammon aries.
 XX
 OS EP386979-A.
 XX
 PN 12-SEP-1990.
 XX
 PD 05-MAR-1990; 90EP-0302322.
 XX
 PF 06-MAR-1989; 89US-0319585.
 XX
 XX (GETH) GENENTECH INC.
 XX
 XX Wood WI, Colosi PC;
 XX
 XX WPI: 1990-276995/37.
 DR P-PSDB; AAR06643.
 XX
 XX DNA sequence encoding growth-hormone-receptor-binding protein -
 PT useful for increase of mammalian meat and milk prodn. and
 PT decrease in fat content.
 XX

PS Claim 5; Page 9; 22pp; English.

XX Receptor binding protein may be used to increase meat and milk

CC production and decrease fat content especially in sheep and goats.

CC Sequence shows an identity of 26% with human growth hormone and 67%

CC to ovine placental lactogen.

XX

SQ Sequence 594 BP; 191 A; 150 C; 125 G; 128 T; 0 other;

Query Match 0.6%; Score 18; DB 11; Length 594;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 ttgaagggaatctggca 517
|||||

Db 54 TTGAAGCGGATCTGGCA 37

RESULT 33

AAF07589

XX AAF07589 standard; cDNA; 652 BP.

AC AAF07589;

DT 13-MAR-2001 (first entry)

XX

DE Fusarium venenatum EST SEQ ID NO:112.

XX

KM Multiple gene expression; filamentous fungal cell; EST;

KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KM Aspergillus oryzae; Trichoderma reesei; Identification; recombination;

KM culture condition; environmental stress; spore morphogenesis;

KM metabolic pathway engineering; catabolic pathway engineering; ss.

XX

OS Fusarium venenatum.

XX

PM WO200056762-A2.

XX

PD 28-SEP-2000.

XX

PF 22-MAR-2000; 2000WO-US07781.

XX

PR 22-MAR-1999; 99US-0273623.

XX

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

XX

PI Berta RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX

DR WPI; 2000-594572/56.

XX

PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a

PT substrate of expressed sequence tags -

XX

PS Claim 86; Page 419; 3161pp; English.

XX

CC The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to

CC expression of the same genes in one or more second filamentous fungal

CC cells. The method uses fluorescence-labeled nucleic acids isolated from

CC the FF cells and a substrate for expressing differential expression of genes

CC are used in the methods for monitoring differential expression of genes

CC in a first filamentous fungal (FF) cell relative to expression of the

CC same genes in one or more second filamentous fungal cells. Monitoring

CC the global expression of genes from FF cells allows the production

CC potential of the microorganisms to be improved. New genes may be

CC discovered, possible functions of unknown open reading frames can be

CC identified and gene copy number variation and stability can be

CC monitored. The expression of genes can be used to study how FF cells

CC adapt to changes in culture conditions, environmental stress, spore

CC morphogenesis, recombination, metabolic or catabolic pathway

CC engineering. Using ESTs provides several advantages over genomic or

CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the

CC microarrays based on function of the gene products to facilitate

CC analysis of the results. AAF07478 to AAF11247 represents ESTs from

CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus

CC niger; AAF11854 to AAF14678 represents ESTs from Aspergillus oryzae; and

CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are

CC all specifically claimed in the present invention.

XX

SQ Sequence 652 BP; 142 A; 185 C; 201 G; 121 T; 3 other;

Query Match 0.6%; Score 18; DB 21; Length 652;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2092 gctggtgctgctggaagc 2109
|||||

Db 578 gctggtgctgctggaagc 595

RESULT 34

AAC44321/C

ID AAC44321 standard; DNA; 682 BP.

XX

AC AAC44321;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 42431.

XX

KM Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152362.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159299.
PR 13-OCT-1999; 99US-0159299.
PR 13-OCT-1999; 99US-0159299.
PR 13-OCT-1999; 99US-0159299.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 0.6%; Score 18; DB 21; Length 682;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 428 accacaccatcacgcg 445
|||||
DB 517 ACCACACCATCACCGC 500

RESULT 35

AAH05406/c
ID AAH05406 standard; CDNA; 814 BP.

AC AAH05406;

DT 26-JUN-2001 (first entry)

DE Human CDNA clone (5'-primer) SEQ ID NO:2241.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PE 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

PT full-length cDNAs -

PS Claim 1; SEQ ID 2241; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC *detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs, easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 814 BP; 155 A; 295 C; 222 G; 136 T; 6 other;

Query Match 0.6%; Score 18; DB 22; Length 814;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 869 ttggtgccttgcctgctg 886
|||||
DB 786 TTGgtGCCTtGCtGCTG 769

RESULT 36

AAAX20212
ID AAAX20212 standard; DNA; 888 BP.

AC AAAX20212;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis gene EF110.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

KW detection; attenuation; antigenic; ss.

OS Enterococcus faecalis.

PN WO9850554-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US089959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromocky A, Kunsch CA;

PI P-PSDB; AAY00222.

PT New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection

PS Claim 1; Page 212; 301pp; English.

XX The present sequence represents a gene isolated from
CC Enterococcus faecalis. The present invention describes genes, proteins
CC and antigenic polypeptides isolated from E. faecalis. The proteins can
CC be used in vaccines for preventing or attenuating an infection caused
CC by a member of the Enterococcus genus in an animal. They can also be
CC used for detecting Enterococcus antibodies in a sample. The nucleotide
CC sequences can be used for detecting Enterococcus nucleic acids.
CC Products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of E. faecalis protein
CC activity.

XX Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 888;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2492 gcttctcgtcaatgcag 2509
 |||||
 DB 110 gcttctcgtcaatgcag 127

RESULT 37
 AAH67086
 ID AAH67086 standard; DNA; 960 BP.
 XX
 AC AAH67086;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 2121.
 XX
 KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 DR P-PSDB; AAG91867.
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX
 PS Claim 8; SEQ ID NO: 2121; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 960 BP; 206 A; 235 C; 271 G; 248 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 960;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3005 ttgtcgggtgcaattg 3022
 |||||
 DB 394 ttgtcgggtgcaattg 411

RESULT 38
 AAQ05870/C

ID AAQ05870 standard; cDNA; 992 BP.
 XX
 AC AAQ05870;
 XX
 DT 07-JAN-1991 (first entry)
 XX
 DE Sequence encoding mammalian growth hormone receptor binding protein.
 XX
 KM Ovine placental lactogen; ds;
 XX
 OS Ovis ammon aries.
 XX
 FH Key Location/Qualifiers
 FT CDS 69..776
 FT /*tag= a
 XX
 PN EP386979-A.
 XX
 PD 12-SEP-1990.
 XX
 PE 05-MAR-1990; 90EP-0302322.
 XX
 PR 06-MAR-1989; 89US-0319585.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Wood WL, Colosi PC;
 XX
 DR WPI; 1990-276995/37.
 DR P-PSDB; AAR06643.
 XX
 PT DNA sequence encoding growth-hormone-receptor-binding protein -
 PT useful for increase of mammalian meat and milk prodn. and
 PT decrease in fat content.
 XX
 PS Claim 14; Fig 3; 22pp; English.
 XX
 CC Receptor binding protein may be used to increase meat and milk
 CC production and decrease fat content especially in sheep and goats.
 CC Sequence shows an identity of 26% with human growth hormone and 67%
 CC to ovine placental lactogen.
 CC Sequence was obtained using probe described in AAQ05869.
 XX
 SQ Sequence 992 BP; 308 A; 237 C; 195 G; 252 T; 0 other;

Query Match 0.6%; Score 18; DB 11; Length 992;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 500 ttgaagggaatcgtgca 517
 |||||
 DB 236 TTGAAGGGGAATCTGCA 219

RESULT 39
 AAZ52527/C
 ID AAZ52527 standard; cDNA; 1034 BP.
 XX
 AC AAZ52527;
 XX
 DT 29-FEB-2000 (first entry)
 XX
 DE Human secreted protein clone ya66_1 nucleotide sequence SEQ ID NO:105.
 XX
 KM Human; secreted protein; immunostimulatory; haemostatic; cytokine;
 KM proliferative; differentiative; chemotactic; chemokinetic; vaccine;
 KM thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
 KM gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9958642-A2.

XX 18-NOV-1999.
PD 14-MAY-1999; 99MO-US10843.
XX 14-MAY-1998; 98US-0085472.
PR 17-AUG-1998; 98US-0096824.
PR 11-SEP-1998; 98US-0099843.
PR 11-SEP-1998; 98US-0099950.
PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 99US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX (GENY) GENETICS INST INC.
PA
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ;
XX
DR WPI; 2000-053095/04.
DR P-PSDB; AAY73442.
XX
PT Novel polynucleotides and proteins having biological activities which
PT make them suitable for treating, preventing or ameliorating medical
PT conditions in humans or animals -
XX
PS Claim 114; Page 653; 730pp; English.
XX
CC The present invention describes human secreted proteins encoded by
CC polynucleotides obtained from adult testes, foetal kidney, adult thymus
CC brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC CDNA libraries. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Suggested activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy. Therapeutic compositions
CC are also presently valuable for veterinary applications. AA252475 to
CC AA252581 encode human secreted proteins, and AAY73390 to AAY73500
CC represent human secreted proteins, given in the present invention.
XX
SQ Sequence 1034 BP; 305 A; 228 C; 222 G; 276 T; 3 other;
XX
Query Match 0.6%; Score 18; DB 21; Length 1034;
Best Local Similarity 100.0%; Pred. No. 14e-02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 415 ctgtgaggaagaccaa 432
|||
Db 168 CTGTGAGGAAGACCAA 151
XX
RESULT 40
AAE71387
ID AAE71387 standard; DNA; 1083 BP.
XX
AC AAE71387;
XX
DT 30-APR-2001 (First entry)
DE .Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:55.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogetic amino acid;
KW nonproteinogetic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diptheriae; evolutionary study; ds.
XX
XX Corynebacterium glutamicum.
XX
PN MO200100844-A2.
PD
XX
PD 04-JAN-2001.
XX
XX 23-JUN-2000; 2000MO-IB00943.
XX
XX 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-015172.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
XX (BADL) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habernauer G;
PI
DR WPI; 2001-061975/07.
DR P-PSDB; AAB79270.
XX
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 3; Page 217-219; 1246pp; English.
XX
XX AAE71360 to AAE71750 encode the Corynebacterium glutamicum sugar
XX metabolism and oxidative phosphorylation (SMP) proteins given in
XX AAB79243 to AAB 79633 which are involved in carbon metabolism and
XX energy production. The C. glutamicum SMP gene can be used in vectors
XX (II) for expression in host cells and production or modulation of
XX production of fine chemicals, such as, an organic acid, a proteinogetic
XX or nonproteinogetic amino acid (preferred), a purine or pyrimidine base,
XX a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
XX acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
XX cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
XX (III) encoded by them are used for diagnosing the presence or activity of
XX Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
XX containing them are used to map genomes of organisms related to

CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).

XX
 SQ Sequence 1083 BP; 241 A; 261 C; 308 G; 273 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 1083;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3005 ttgttggtgcaattg 3022
 ||||||||||||||||

DB 494 ttgttggtgcaattg 511

RESULT 41
 AA242236
 ID AA242236 standard; cDNA; 1205 BP.

XX
 AC AA242236;

XX
 DT 31-JAN-2000 (first entry)

XX
 DE Human normal bladder tissue cDNA derived EST 115.

XX
 KM Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
 KM cancer; gene therapy; ss.

XX
 OS Homo sapiens.

XX
 PN DE19818620-A1.

XX
 PD 28-OCT-1999.

XX
 PF 21-APR-1998; 98DE-1018620.

XX
 PR 21-APR-1998; 98DE-1018620.

XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MDH.

XX
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX
 DR WPI: 1999-602416/52.

XX
 DR P-PSDB; AA160559, AA160560, AA160561.

XX
 PR New polypeptides and their nucleic acids, useful for treatment of
 PR bladder tumour and identification of therapeutic agents -
 XX
 PS Claim 3; Page 235; 366pp; German.

XX
 CC This invention describes novel polypeptide fragment sequences (I) and
 CC their encoding nucleic acids (II) which are highly expressed in normal
 CC bladder tissue and have cytostatic activity. (II) are used for
 CC recombinant expression of (I) and to isolate complete genes. (I) are
 CC used to identify agents suitable for the treatment of bladder tumours,
 CC to directly treat this form of cancer (including expression from gene
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)
 CC is also used for the generation of specific antibodies. (II) are
 CC identified by assembling ESTs (expressed sequence tags) from a
 CC particular tissue type before comparison of expression patterns. This
 CC allows a significantly longer fragment of the gene to be revealed, and
 CC therefore reduces the number of failures because of ESTs from different
 CC libraries representing different parts of the same unknown gene
 CC disrupting the estimated frequency of occurrence in a particular tissue.
 CC AA242122-242248 represent EST fragments derived from a human normal
 CC bladder tissue cDNA library which encode the protein fragments
 CC represented in AA160329-Y60591.

XX
 SQ Sequence 1205 BP; 274 A; 349 C; 338 G; 244 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 1205;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1981 actctgaggtcacactc 1998
 ||||||||||||||||

DB 714 actctgaggtcacactc 731

RESULT 42
 AAN80316
 ID AAN80316 standard; DNA; 1250 BP.

XX
 AC AAN80316;

XX
 DT 10-JAN-1991 (first entry)

XX
 DE Transcription control region of gene encoding serine protease

DE (SP) of human myeloid cell origin.

XX
 KM Serine protease; myeloid cell; intravascular coagulation treatment; ss.

XX
 OS Homo sapiens.

XX
 PN W08806621-A.

XX
 PD 07-SEP-1988.

XX
 PF 26-FEB-1988; 88WO-JP00205.

XX
 PR 09-SEP-1987; 87JP-0225540.

XX
 PR 05-MAR-1987; 87JP-0050676.

XX
 PA (TORA) TORAY IND INC (AOKI/).

XX
 PI Aoki Y, Okano K, Naruto M, Shimizu H, Nakamura H;

XX
 DR WPI: 1988-271161/38.

XX
 PR Serine protease of human myeloid cell origin -

XX
 PR useful in disseminated intravascular coagulation treatment

XX
 PS Disclosure; Fig 15; 52pp; Japanese.

XX
 CC SP DNA is selected and cloned from a human myeloid cell library,
 CC inserted in a suitable vector and expressed in a transformant of a
 CC suitable organism such as E.coli HMS-174 or HB-101 or a yeast. It is new
 CC so are a SP of human myeloid cell origin; precursors of SP having an
 CC N-terminal splittable or signal peptide; and a transcription-controlling
 CC DNA sequence required for expression of the gene. SP has antithrombotic
 CC activity, esp. useful in disseminated intravascular coagulation. By
 CC expression of the protease gene in a suitable transformant organism SP
 CC can be obtained in large quantity.

XX
 SQ Sequence 1250 BP; 305 A; 340 C; 385 G; 220 T; 0 other;

Query Match 0.6%; Score 18; DB 9; Length 1250;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1046 tgggtgaagtgaattcc 1063
 ||||||||||||||||

DB 43 tgggtgaagtgaattcc 60

RESULT 43
 AAC40012

ID AAC40012 standard; DNA; 1473 BP.

XX
 AC AAC40012;

DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 26727.
DE
XX Hybridisation assay: genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132853.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135569.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.


```

XX Brandt KS, Silver GM, Wisniewski N;
PT
XX
XX WPI: 1998-297929/26.
DR
XX P-PSDB: AAW57852.
XX
PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
XX ectoparasites, particularly on cats and dogs
XX
PS Claim 1; Page 113-115; 230pp; English.
XX
CC This sequence encodes the flea esterase protein, nFE51515 (the
CC complementary strand is shown in AAV40738), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.
XX
SQ Sequence 1515 BP; 504 A; 268 C; 314 G; 428 T; 1 other;

Query Match          0.6%; Score 18; DB 19; Length 1515;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
   |||||||
DB 363 CTCGAATTCAGAAATCC 346

RESULT 46
AAV40738
ID AAV40738 standard; cDNA; 1515 BP.
XX
AC AAV40738;
XX
DT 23-SEP-1998 (first entry)
XX
DE C. felis esterase, nFE51515, coding sequence complementary strand.
XX
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51515; ds.
XX
OS Ctenocephalides felis.
XX
PN WO9821324-A1.
XX
PD 22-MAY-1998.
XX
PF 10-NOV-1997; 97WO-US20598.
XX
PR 12-NOV-1996; 96US-0747221.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Silver GM, Wisniewski N;
DR
XX WPI: 1998-297929/26.
XX
PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous

```

```

PT ectoparasites, particularly on cats and dogs
XX
XX Claim 1; Page 115-116; 230pp; English.
PS
XX
XX This sequence is the complementary strand of the DNA encoding the flea
XX esterase protein, nFE51515 (see AAV40737 for coding strand), of the
XX invention. When administered to animals, the protein induces a
XX protective immune (antibody) response against carboxylesterase (CE), so
XX they, or compositions containing CE proteins, are used, therapeutically
XX or as vaccines, to protect particularly mammals and birds, specifically
XX cats and dogs, against haematophagous ectoparasite infestation (HEP),
XX specifically fleas. More generally the compositions can be used to treat
XX arthropods generally, including pests of agricultural crops, trees,
XX stored goods etc., also those that are vectors of disease. Fragments of
XX the DNA can be used as probes and primers for identification or
XX production of nucleic acid. Antibodies against the protein can be used
XX for passive immunisation; to screen expression libraries; to isolate the
XX protein and to target cytotoxic compounds to HEP. The compounds
XX containing CE are effective against both adult and larval stages; they
XX target CEs, including juvenile hormone, that are involved in
XX development, metamorphosis, feeding, digestion and reproduction.
XX
SQ Sequence 1515 BP; 428 A; 314 C; 268 G; 504 T; 1 other;

Query Match          0.6%; Score 18; DB 19; Length 1515;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
   |||||||
DB 1153 ctccaattcagaatcc 1170

RESULT 47
AAF21032
ID AAF21032 standard; DNA; 1561 BP.
XX
AC AAF21032;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2599.
XX
KW low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antisthmatic; analgesic; hypotensive; cyostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasocostriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI (NYCE/) NYCE J W.
XX
PI Nyce JW;
DR
XX WPI: 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not

```

PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

PS Disclosure: Page 840; 1592pp; English.

CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiallergic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, emphysema, chronic obstructive pulmonary disease (COPD),
 CC hyperlaxation, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

SO Sequence 1561 BP; 354 A; 445 C; 485 G; 277 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 1561;

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1046 tgggtgaaagtgaattcc 1063
 |||||
 DB 142 tgggtgaaagtgaattcc 159

RESULT 48
 ID AAA34910
 ID AAA34910 standard; DNA; 1561 BP.

AC AAA34910;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide SEQ ID NO:2599.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KM phosphorochiolate; impaired respiration; inflammation; allergy;
 KM allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KM antiallergic; antiallergic; cytostatic; analgesic; impaired airway;
 KM lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KM respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KM pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KM cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

PN WO200009525-A2.

PD 24-FEB-2000.

PF 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Myce JW;

DR WPI; 2000-205971/18.

PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -

PS Disclosure: Page 769; 1343pp; English.

CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiallergic, cytostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

SO Sequence 1561 BP; 354 A; 445 C; 485 G; 277 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 1561;

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1046 tgggtgaaagtgaattcc 1063
 |||||
 DB 142 tgggtgaaagtgaattcc 159

RESULT 49

ID AAH13702/c
 ID AAH13702 standard; cDNA; 1579 BP.

AC AAH13702;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:10578.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length CDNA defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length CDNA -
 PS
 PS Claim 8; SEQ ID 10578; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length CDNA defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length CDNA. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length CDNA. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 XX
 SQ Sequence 1579 BP; 468 A; 274 C; 283 G; 554 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 1579;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaattcataaagac 18
 ||||||||||||||||
 DB 1229 ATGAATTTCATPAAAGGAC 1212

RESULT 50
 AAQ12528/c
 ID AAQ12528 standard; DNA; 1611 BP.
 XX
 AC AAQ12528;
 XX
 DT 25-SEP-1991 (first entry)
 XX
 DE Thymidylate phosphorylase.
 XX
 XX
 KM deoxyribonucleic acid; thymidine; dTMP; pyrimidine; ss.
 XX
 OS Bacillus subtilis CMG356 (ATCC 33234).
 XX
 OS
 XX
 FT .Key Location/Qualifiers
 FT CDS 136..857
 FT /tag= a

FT /product= dTMPase
 XX
 XX WO9109130-A.
 XX
 PD 27-JUN-1991.
 XX
 XX
 XX 05-DEC-1990; 90WO-US06993.
 XX
 PR 08-DEC-1989; 89US-0448158.
 XX
 PA (CHEM-) CHEMGEN CORP.
 XX
 PI Mc Dandliss RL, Anderson DM;
 DR WPI; 1991-208156/28.
 DR P-PSDB; AAR12555.
 XX
 PT Microorganism contg. deoxyribonucleic acid - encoding enzyme
 PT causing accumulation of pyrimidine deoxyribonucleoside in
 PT recoverable amts.
 PS
 PS Disclosure; Fig 7(a-c); 79pp; English.
 XX
 CC The sequence, comprising flanking DNA sequences obt'd. from a pScl
 CC bacteriophage, encodes an enzyme capable of converting a thymidine
 CC deoxyribonucleoside monophosphate to a thymidine deoxyribonucleoside.
 CC The sequence may be used together with metabolic mutations of
 CC heterologous DNA, encoding metabolic enzymes, to engineer cultured
 CC cells to express thymidine deoxyribonucleosides (TdNs) in recoverable
 CC amts., providing a fermentation source of TdNs.
 CC See also AAQ12526-27.
 CC
 XX
 SQ Sequence 1611 BP; 683 A; 171 C; 244 G; 513 T; 0 other;

Query Match 0.6%; Score 18; DB 12; Length 1611;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 aggaattcaagactgt 325
 ||||||||||||||||
 DB 979 AGGATTTAAAGGACTTGT 962

RESULT 51
 AAA47151/c
 ID AAA47151 standard; DNA; 1611 BP.
 XX
 AC AAA47151;
 XX
 DT 03-OCT-2000 (first entry)
 XX
 DE DNA encoding a serine protease inhibitor protein.
 XX
 XX
 KM Serine protease inhibitor; green-lipped mussel; anti-thrombin;
 KM divalent metal cation binding activity; dietary supplement;
 KM anticoagulant; ss.
 XX
 OS Perna canaliculus.
 XX
 OS
 XX
 FT .Key Location/Qualifiers
 FT CDS 1..1494
 FT /tag= a
 FT /product= "serine protease inhibitor"
 FT polyA_site 1557..1563
 FT /tag= b
 XX
 PN WO200039165-A1.
 XX
 PD 06-JUL-2000.
 XX
 PF 23-DEC-1999; 99WO-NZ00227.

PR 23-DEC-1998; 98NZ-033568.
 PR 23-JUL-1999; 99NZ-0336906.
 XX
 XX
 PA. (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
 XX
 PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
 XX
 DR WPI; 2000-452375/39.
 XX P-PSDB; AAY93750.
 XX
 PT New Perna canaliculus serine protease inhibitor protein exhibiting
 PT anti-thrombin activity and divalent metal cation binding activity,
 PT useful as an anticoagulant agent and as a dietary supplement -
 XX
 PS Claim 11; Page 10-11; 44pp; English.
 XX
 CC The present sequence encodes a serine protease inhibitor
 CC protein. The protein is isolated from the green-lipped mussel
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
 CC and divalent metal cation binding activity. The serine protease
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
 CC The protein, and its fragments, are useful in medicaments, in food,
 CC as dietary supplements or as bioremediation agents. In the dietary
 CC supplements, the protein is associated with or bound to at least one
 CC divalent cation (such as calcium, magnesium or zinc) of dietary
 CC significance. The proteins or their fragments are also useful as
 CC anticoagulant agents.
 XX
 SQ Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other:
 Query Match 0.6%; Score 18; DB 21; Length 1611;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1016 tctttgtggtacacatt 1033
 Db 767 TCTTTGTGTCATCACTT 750
 |||||
 RESULT 52
 AA15908/C
 ID AA15908 standard; cDNA; 1619 BP.
 XX
 AC AA15908;
 XX
 DT 12-JUN-2000 (first entry)
 XX
 DE Human protein clone HP10195 full length coding sequence.
 XX
 KW Human protein; hydrophobic domain; nutritional source; hematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200005367-A2.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-JP03929.
 XX
 PR 24-JUL-1998; 98JP-0208820.
 PR 07-AUG-1998; 98JP-0224105.
 PR 25-AUG-1998; 98JP-0238116.

PR 09-SEP-1998; 98JP-0254736.
 PR 29-SEP-1998; 98JP-0275505.
 XX
 XX
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 XX
 DR WPI; 2000-182694/16.
 XX P-PSDB; AAY94850.
 XX
 PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX
 PS Claim 4; Page 192-194; 351pp; English.
 XX
 CC This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of hematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.
 XX
 SQ Sequence 1619 BP; 444 A; 359 C; 363 G; 453 T; 0 other:
 Query Match 0.6%; Score 18; DB 21; Length 1619;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgaattcacaagac 18
 Db 1277 ATGAATTTCATTAAGAC 1260
 |||||
 RESULT 53
 AA140760/C
 ID AA140760 standard; cDNA; 1650 BP.
 XX
 AC AA140760;
 XX
 DT 23-SEP-1998 (first entry)
 XX
 DE C. felis esterase, nFE51650, coding sequence.

```

XX Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51650; ds.
XX
OS Ctenocephalides felis.
XX
PN W09821324-A1.
XX
PD 22-MAY-1998.
XX
PF 10-NOV-1997; 97WO-US20598.
XX
PR 12-NOV-1996; 96US-0747221.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Silver GM, Wisniewski N;
XX
DR WPI; 1998-297929/26.
DR P-PSDB; AAW57865.
XX
PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
XX
PS Claim 1; Page 188-191; 230pp; English.
XX
CC This sequence encodes the flea esterase protein, nFE51650 (the
CC complementary strand is shown in AAV40761), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.
XX
SQ Sequence 1650 BP; 542 A; 284 C; 341 G; 482 T; 1 other;
XX
Query Match 0.6%; Score 18; DB 19; Length 1650;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 967 ctccaattcagaatcc 984
DB 498 CTCCAATTTCAGAAATCC 481
XX
RESULT 54
AAV40761
ID AAV40761 standard; cDNA; 1650 BP.
XX
AC AAV40761;
XX
DT 23-SEP-1998 (first entry)
XX
DE C. felis esterase, nFE51650, coding sequence complementary strand.
XX
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51650; ds.
XX
OS Ctenocephalides felis.
XX
PN W09821324-A1.
XX

```

```

XX
PD 22-MAY-1998.
XX
PF 10-NOV-1997; 97WO-US20598.
XX
PR 12-NOV-1996; 96US-0747221.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Silver GM, Wisniewski N;
XX
DR WPI; 1998-297929/26.
XX
PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
XX
PS Claim 1; Page 191-192; 230pp; English.
XX
CC This sequence is the complementary strand of the DNA encoding the flea
CC esterase protein, nFE51650 (see AAV40760 for coding strand), of the
CC invention. When administered to animals, the protein induces a
CC protective immune (antibody) response against carboxylesterase (CE), so
CC they, or compositions containing CE proteins, are used, therapeutically
CC or as vaccines, to protect particularly mammals and birds, specifically
CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.
XX
SQ Sequence 1650 BP; 482 A; 341 C; 284 G; 542 T; 1 other;
XX
Query Match 0.6%; Score 18; DB 19; Length 1650;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 967 ctccaattcagaatcc 984
DB 1153 ctccaattcagaatcc 1170
XX
RESULT 55
AAV71287/C
ID AAV71287 standard; DNA; 1723 BP.
XX
AC AAV71287;
XX
DT 12-AUG-1999 (first entry)
XX
DE Human vesicular binding protein nucleotide sequence.
XX
KW VSBP; fibroblast; Aplysia; VAMP; vesicle associated membrane protein;
KW synapobrevin binding protein; cancer; inflammation; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 568..1278
FT /*tag= a
FT /*product= "VWBP"
XX
PN W09851797-A1.
XX
PD 19-NOV-1998.
XX

```

PF 15-MAY-1998; 98WO-US10225.
 XX
 PR 15-MAY-1997; 97US-0857213.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Hawkins PR, Murry LE;
 DR WPI: 1999-045229/04.
 XX P-PSDB: AAM86813.
 PT Human vesicle binding protein - useful for the diagnosis, treatment
 PT and prevention of cell proliferation disorders and inflammation
 PS
 PS Claim 5: Fig 1A-D; 58pp: English.
 XX
 CC The present sequence represents a nucleic acid encoding a purified human
 CC vesicle binding protein (designated MVBP). This was obtained by standard
 CC cloning and screening procedure, where it was first isolated in incyte
 CC clone 148415 from a normal fibroblast library (FIRNG701). A comparison
 CC of the MVBP with the Aplysia vesicle-associated membrane protein
 CC (VAMP)/synaptobrevin binding protein showed that they shared 44%
 CC identity. MVBP and its products can be used to diagnose, treat or prevent
 CC disorders of cell proliferation (e.g. cancer) or inflammation. Sequences
 CC complementary to MVBP can be used to detect MVBP in samples. Ads are used
 CC to diagnose conditions associated with MVBP or in assays to monitor
 CC patients being treated with MVBP or agonists/antagonists.
 CC
 SQ Sequence 1723 BP; 548 A; 356 C; 325 G; 468 T; 26 other;

Query Match 0.6%; Score 18; DB 20; Length 1723;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgaattcataaagac 18
 ||||||||||||
 Db 1540 ATGAATTTCATTAAGAC 1523

RESULT 56
 AA067601
 ID AA067601 standard; cDNA: 1800 BP.
 XX
 AC AA067601;
 XX
 DT 20-NOV-1994 (first entry)
 XX
 DE Retinoblastoma protein Ap4.
 XX
 KW Retinoblastoma-associated polypeptide; RAP; Ap12; Ap4;
 KW transcription factor; BP2; cell cycle; ss.
 OS Homo sapiens.
 XX
 PN MO9412521-A.
 XX
 PD 09-JUN-1994.
 XX
 PF 19-NOV-1993; 93WO-US11310.
 XX
 PR 20-NOV-1992; 92US-0979156.
 XX
 PA (TEKA) UNIT TEXAS SYSTEM.
 XX
 PI Lee W, Shan B;
 XX
 DR WPI: 1994-200185/24.
 XX
 PT Nucleic acid encoding retinoblastoma-associated polypeptide(s) -
 PT used for producing prods. for use in studying cell cycling and
 PT deregulated cell growth.
 XX

PS Disclosure: Page 37-38; 77pp: English.
 XX
 CC cDNA encoding a retinoblastoma (RB) associated protein (RAP), Ap12,
 CC was cloned by direct screening of cDNA expression libraries using
 CC purified RB protein as probe. Ap12 showed transcription factor E2F
 CC and RB binding activities. The 5' and 3' cDNA sequences were
 CC determined for Ap2 (AA067395, AA067596), Ap8 (AA067597, AA067598) and
 CC Ap15 (AA067599, AA067600), and full-length encoding sequences for Ap4
 CC (AA067601) and Ap10 (AA067601), which also encoded RB-associated
 CC proteins.
 XX
 SQ Sequence 1800 BP; 747 A; 333 C; 372 G; 348 T; 0 other;

Query Match 0.6%; Score 18; DB 15; Length 1800;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2796 aaagaccctcgaataa 2813
 ||||||||||||
 Db 307 aaagaccctcgaataa 324

RESULT 57
 AAF25374
 ID AAF25374 standard; cDNA: 1878 BP.
 XX
 AC AAF25374;
 XX
 DT 15-MAY-2001 (first entry)
 XX

Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
 XX
 DE
 XX
 KW SGS3 gene: post-transcriptional inactivation; RNA degradation;
 KW viral resistance; resistance; fatty acid content; protein content; ss.
 XX
 OS Arabidopsis thaliana.
 XX

Key Location/Qualifiers
 FT 1..1878
 FT CDS /*tag= a
 FT /*product= "SGS3"

WO200105951-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000WO-FR02052.
 XX
 PR 16-JUL-1999; 99FR-0009417.
 PR 26-JAN-2000; 2000FR-0001006.
 XX
 PA (AVET) AVENTIS CROSCIENCE SA.
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 XX
 PI Beclin C, Elmayan T, Vaucheret H;
 XX
 DR WPI: 2001-159529/16.
 DR P-PSDB: AAB31798.
 XX
 PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression
 XX
 PS Claim 1: Page 32-35; 36pp: French.
 XX

The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.
 CC The SGS3 gene is essential for post-transcriptional inactivation
 CC (degradation of RNA) and for resistance to viruses. Overexpression
 CC of SGS3 results in plants with increased resistance to viruses,
 CC while inactivation of SGS3 in transgenic plants (e.g. by expressing
 CC antisense RNA, by mutation or by homologous recombination) increases
 CC the level of the transgene product. This product may e.g. impart

CC resistance (to herbicide, insects or pathogens), alter contents
 CC of essential fatty acids or proteins, or is pharmaceutically active,
 CC e.g. an immunoglobulin or interferon.
 XX
 SQ Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 1878;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 985 agtgcctgaggatctc 1002
 |||
 Db 1388 agtgcctgaggatctc 1405

RESULT 58
 AAV82522
 ID AAV82522 standard; DNA; 1926 BP.
 XX
 AC AAV82522;
 XX
 DT 17-MAR-1999 (first entry)
 XX

DE Sphingomonas capsulata mature aminopeptidase I encoding DNA.
 XX
 KM Aminopeptidase; protein hydrolyase; glycine releasing; protease;
 KM protease; material; flavour; food; baking; animal feed additive;
 KM palatability; hydrolysis; solubility; emulsifying; foaming; aroma;
 KM mouthfeel; crust colour; ss.
 XX

OS Sphingomonas capsulata.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1491
 FT CDS /*tag= a

W09851163-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98MO-US09998.
 XX

PR 16-DEC-1997; 97US-0069719.
 PR 16-MAY-1997; 97US-0857886.
 PR 20-OCT-1997; 97US-0062893.
 PR 16-DEC-1997; 97DK-0001465.
 XX

PA (ASAH) ASAH CHEM IND CO LTD.
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO-NORDISK AS.
 XX

PI Blinovsky A, Brown K, Byun T, Fujii M, Golightly E;
 PI Kofod LV, Marumotac, Mathiasen TE;
 XX
 DR WPI; 1999-045177/04.
 DR P-PSDB; AAW89587.
 XX

PT Production of protein hydrolylate - using protease and enzyme that
 PT releases glycine, useful as flavour improvers in foods and animal
 PT feed additives
 XX

PS Claim 7; Page 61; 84pp; English.
 XX

CC A method has been developed for the production of protein hydrolysates
 CC (PH) comprising reacting a protein with: (i) at least one polypeptide
 CC with gly-releasing activity; and (ii) at least one other protease so
 CC that the amount of Gly produced is greater than when the protease is
 CC used alone. PH are used to improve flavour of foods (e.g. baked goods)
 CC and as animal feed additives. PH are preferably also enriched in Glu
 CC (free and/or peptide bound), so have improved flavour and palatability.
 CC Addition of a polypeptide with gly-releasing activity increases the
 CC degree of hydrolysis (or reduces the amount of enzyme needed) and

CC hydrolysates have better solubility, and emulsifying and foaming
 CC properties. Baked goods containing them have improved aroma, mouthfeel
 CC and crust colour. The present sequence encodes Sphingomonas capsulata
 CC aminopeptidase I, which is used in the method of the invention.
 XX
 SQ Sequence 1926 BP; 362 A; 633 C; 628 G; 303 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 1926;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2092 gctggtgtgctggaagc 2109
 |||
 Db 1242 gctggtgtgctggaagc 1259

RESULT 59
 AA277504
 ID AA277504 standard; cDNA; 1962 BP.
 XX
 AC AA277504;
 XX
 DT 10-APR-2000 (first entry)
 XX

DE Human ovarian tumor cDNA library derived EST fragment 55.
 XX

KM Expressed sequence tag; EST; human; ovarian tumor; anticancer;
 KM gene therapy; treatment; ss.
 XX

OS Homo sapiens.
 XX

PN DE19817557-A1.
 XX

PD 21-OCT-1999.
 XX

PF 09-APR-1998; 98DE-1017557.
 XX

PR 09-APR-1998; 98DE-1017557.
 XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 PI WPI; 1999-591920/51.
 DR P-PSDB; AAV76633, AAV76634, AAV76635.
 DR
 XX

PT New nucleic acid sequences expressed in ovarian, and some other, cancer
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and
 PT identification of therapeutic agents
 XX

PS Claim 3; Page 182-183; 310pp; German.
 XX

CC This invention describes novel nucleic acid (cDNA) sequences (A) which
 CC have anticancer activity and are highly expressed in ovarian tumor
 CC tissue (and some also in testis and breast cancer tissue). The products
 CC of the invention can be used for gene therapy. (A) are used (i) for
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete
 CC genes. (B) are used (i) to identify agents suitable for treatment of
 CC ovarian cancer; (ii) directly for treating this form of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AA277450-277572 represent the human
 CC ovarian tumor cDNA library derived EST fragments described in the method
 CC of the invention and encode the protein fragments represented in
 CC AAV7505-Y76638.
 XX

Sequence 1962 BP; 596 A; 357 C; 389 G; 620 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 1962;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaattcataaaggac 18
|||||
Db 368 atgaattcataaaggac 385

RESULT 60
AAV40735/C
ID AAV40735 standard; cDNA; 1982 BP.

XX AAV40735;

DT 23-SEP-1998 (first entry)

DE C. felis esterase, nFE51982, coding sequence.

XX Esterase; flea; protective immune response; carboxylesterase; arthropod;

KW haematophagous ectoparasite infestation; nFE51982; ds.

OS Clenocephalides felis.

FT Key Location/Qualifiers

FT CDS 31..1520

FT /*tag= a

PN W09821324-A1.

PD 22-MAY-1998.

XX 10-NOV-1997; 97WO-US20598.

XX 12-NOV-1996; 96US-0747221.

XX (HESK-) HESKA CORP.

PI Brandt KS, Silver GM, Wisniewski N;

DR WPI: 1998-297929/26.

DR P-PSDB; AAM57851.

XX New nucleic acid encoding carboxylesterase(s) from fleas - useful,

PT e.g. in vaccines, for preventing infestation by haematophagous

PT ectoparasites, particularly on cats and dogs

XX Claim 1; Page 106-109; 230pp; English.

XX This sequence encodes the flea esterase protein, nFE51982 (the

CC complementary strand is shown in AAV40735), of the invention. When

CC administered to animals, the protein induces a protective immune

CC (antibody) response against carboxylesterase (CE), so they, or

CC compositions containing CE proteins, are used, therapeutically or as

CC vaccines, to protect particularly mammals and birds, specifically cats

CC and dogs, against haematophagous ectoparasite infestation (HEP),

CC specifically fleas. More generally the compositions can be used to treat

CC arthropods generally, including pests of agricultural crops, trees,

CC stored goods etc., also those that are vectors of disease. Fragments of

CC the DNA can be used as probes and primers for identification or

CC production of nucleic acid. Antibodies against the protein can be used

CC for passive immunisation; to screen expression libraries; to isolate the

CC protein and to target cytotoxic compounds to HEP. The compounds

CC containing CE are effective against both adult and larval stages; they

CC target CEs, including juvenile hormone, that are involved in

CC development, metamorphosis, feeding, digestion and reproduction.

XX Sequence 1982 BP; 691 A; 304 C; 357 G; 629 T; 1 other;

Query Match 0.6%; Score 18; DB 19; Length 1982;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcgaatcc 984
|||||
Db 365 ctccaattcgaatcc 348

RESULT 61
AAV40736
ID AAV40736 standard; cDNA; 1982 BP.

XX AAV40736;

DT 23-SEP-1998 (first entry)

DE C. felis esterase, nFE51982, coding sequence complementary strand.

XX Esterase; flea; protective immune response; carboxylesterase; arthropod;

KW haematophagous ectoparasite infestation; nFE51982; ds.

OS Clenocephalides felis.

PN W09821324-A1.

PD 22-MAY-1998.

XX 10-NOV-1997; 97WO-US20598.

XX 12-NOV-1996; 96US-0747221.

XX (HESK-) HESKA CORP.

PI Brandt KS, Silver GM, Wisniewski N;

DR WPI: 1998-297929/26.

XX New nucleic acid encoding carboxylesterase(s) from fleas - useful,

PT e.g. in vaccines, for preventing infestation by haematophagous

PT ectoparasites, particularly on cats and dogs

XX Claim 1; Page 111-112; 230pp; English.

XX This sequence is the complementary strand of the DNA encoding the flea

CC esterase protein, nFE51982 (see AAV40735 for coding strand), of the

CC invention. When administered to animals, the protein induces a

CC protective immune (antibody) response against carboxylesterase (CE), so

CC they, or compositions containing CE proteins, are used, therapeutically

CC or as vaccines, to protect particularly mammals and birds, specifically

CC cats and dogs, against haematophagous ectoparasite infestation (HEP),

CC specifically fleas. More generally the compositions can be used to treat

CC arthropods generally, including pests of agricultural crops, trees,

CC stored goods etc., also those that are vectors of disease. Fragments of

CC the DNA can be used as probes and primers for identification or

CC production of nucleic acid. Antibodies against the protein can be used

CC for passive immunisation; to screen expression libraries; to isolate the

CC protein and to target cytotoxic compounds to HEP. The compounds

CC containing CE are effective against both adult and larval stages; they

CC target CEs, including juvenile hormone, that are involved in

CC development, metamorphosis, feeding, digestion and reproduction.

XX Sequence 1982 BP; 629 A; 357 C; 304 G; 691 T; 1 other;

Query Match 0.6%; Score 18; DB 19; Length 1982;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcgaatcc 984
|||||
Db 1618 ctccaattcgaatcc 1635

RESULT 62
AAV40758/c
ID AAV40758 standard; cDNA: 2144 BP.
XX
XX
AC AAV40758;
XX
XX 23-SEP-1998 (first entry)
DT
DE C. felis esterase, nFE52144, coding sequence.
XX
XX
KM Esterase: flea; protective immune response: carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE52144; ds.
XX
XX
OS Ctenocephalides felis.
XX
XX
FH Key Location/Qualifiers
FT CDS 30..1682
FT /tag= a
XX
XX
PN W09821324-A1.
XX
XX 22-MAY-1998.
PD
XX
XX 10-NOV-1997; 97WO-US20598.
PF
XX
XX 12-NOV-1996; 96US-0747221.
PR
XX
XX (HESK-) HESKA CORP.
PA
XX
XX Brandt KS, Silver GM, Wisniewski N;
PI
XX
XX WPI: 1998-297929/26.
DR
XX
XX P-PSDB; AAW57864.
XX
XX
PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
XX
XX
PS Claim 1; Page 181-184; 230pp; English.
XX
XX
CC This sequence encodes the flea esterase protein, nFE52144 (the
CC complementary strand is shown in AAV40759), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically
CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.
XX
XX
SO Sequence 2144 BP; 739 A; 326 C; 389 G; 689 T; 1 other;

Query Match 0.6%; Score 18; DB 19; Length 2144;
Best local Similarity 100.0%; Pred. NO. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
IIIIIIIIIIIIIIIIIIII
DB 527 CTCCAATTTCAGAAATCC 510

RESULT 63
AAV40759
ID AAV40759 standard; cDNA: 2144 BP.

XX
AC AAV40759;
XX
XX
DT 23-SEP-1998 (first entry)
XX
XX
DE C. felis esterase, nFE52144, coding sequence complementary strand.
XX
XX
KM Esterase: flea; protective immune response: carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE52144; ds.
XX
XX
OS Ctenocephalides felis.
XX
XX
PN W09821324-A1.
XX
XX 22-MAY-1998.
PD
XX
XX 10-NOV-1997; 97WO-US20598.
PF
XX
XX 12-NOV-1996; 96US-0747221.
PR
XX
XX (HESK-) HESKA CORP.
PA
XX
XX Brandt KS, Silver GM, Wisniewski N;
PI
XX
XX WPI: 1998-297929/26.
DR
XX
XX
PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
XX
XX
PS Claim 1; Page 187-188; 230pp; English.
XX
XX
CC This sequence is the complementary strand of the DNA encoding the flea
CC esterase protein, nFE52144 (see AAV40758 for coding strand), of the
CC invention. When administered to animals, the protein induces a
CC protective immune (antibody) response against carboxylesterase (CE), so
CC they, or compositions containing CE proteins, are used, therapeutically
CC or as vaccines, to protect particularly mammals and birds, specifically
CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.
XX
XX
SO Sequence 2144 BP; 689 A; 389 C; 326 G; 739 T; 1 other;

Query Match 0.6%; Score 18; DB 19; Length 2144;
Best local Similarity 100.0%; Pred. NO. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 967 ctccaattcagaatcc 984
IIIIIIIIIIIIIIIIIIII
DB 1618 ctccaattcagaatcc 1635

RESULT 64
AAZ96326/c
ID AAZ96326 standard; DNA: 2299 BP.
XX
XX
AC AAZ96326;
XX
XX
DT 10-APR-2000 (first entry)
XX
XX
DE S. pneumoniae derived DNA from ORF #154.
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;

KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX Streptococcus pneumoniae.
OS
XX MO9806734-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US14436.
XX
XX 16-AUG-1996; 96US-0024022.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
XX WPI: 1998-159452/14.
DR
DR P-PSDB: AAY85980, AAY85981.
XX
XX Streptococcus pneumoniae proteins and related DNA - useful for
PT screening compounds for antibacterial activity
PS
XX Claim 4; Page 188-189; 640pp; English.
XX
XX This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AA296173-296494) and their encoded proteins (see
CC AA85792-186182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.
XX
XX Sequence 2299 BP; 585 A; 514 C; 492 G; 708 T; 0 other;
SQ

Query Match 0.6%; Score 18; DB 19; Length 2299;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2730 caagcttggtgaa 2747
|||||
DB 429 CAAGCTTGCTTGAAAA 412

RESULT 65
AAV42995/C
ID AAV42995 standard; DNA: 2299 BP.
XX
XX AAV42995;
AC
XX 09-NOV-1998 (first entry)
DT
XX
XX Streptococcus pneumoniae polypeptide coding region.
DE
XX
XX Polypeptide: ORF; open reading frame; infection; bacterial;
KW streptococcal; bacteremia; diagnosis; prophylaxis; ds.
XX
XX Streptococcus pneumoniae.
OS
XX
XX Key location/qualifiers
FH complement (388..1008)
FT /**tag= a
FT /note= "polypeptide"
XX
XX MO9823631-A1.

XX
XX 04-JUN-1998.
PD
XX
XX 24-NOV-1997; 97WO-US21976.
PE
XX
XX 27-NOV-1996; 96US-0031879.
PR
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Reid RH, Zarfos PN;
XX
XX WPI: 1998-322654/28.
DR
DR P-PSDB: AAM62719.
XX
XX Streptococcus pneumoniae polynucleotides - useful for developing
PT products for diagnosis, prevention and treatment of infections e.g.
PT pneumonia, bacteremia, meningitis or endocarditis
PS
XX Claim 1; Page 106; 181pp; English.
XX
XX The sequence is that of a Streptococcal polypeptide coding region.
CC The polypeptide can potentially be used for the diagnosis and
CC prevention of bacterial infections, especially sp infection.
CC It may be used for the treatment of diseases such as otitis media,
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC empyema, endocarditis or infection of the cerebrospinal fluid.
XX
XX Sequence 2299 BP; 585 A; 514 C; 491 G; 708 T; 1 other;
SQ

Query Match 0.6%; Score 18; DB 19; Length 2299;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2730 caagcttggtgaa 2747
|||||
DB 429 CAAGCTTGCTTGAAAA 412

RESULT 66
AAAX84399
ID AAAX84399 standard; DNA: 3000 BP.
XX
XX AAAX84399;
AC
XX 09-SEP-1999 (first entry)
DT
XX
XX S. capsulata IF012533 aminopeptidase coding sequence.
DE
XX
XX Aminopeptidase; amino acid removal; protein hydrolyate production;
KW hydrolysis; flavour development; enzyme deactivation;
KW peptide sequence cleavage; post-translational processing;
KW precursor protein activation; ss.
XX
XX Sphingomonas capsulata.
OS
XX
XX MO9931226-A1.
PN
XX
XX 24-JUN-1999.
PD
XX
XX 13-NOV-1998; 98WO-DK00495.
PE
XX
XX 15-MAY-1998; 98DK-0000670.
PR 16-DEC-1997; 97DK-0001465.
PR 16-DEC-1997; 97US-0069719.
XX
XX (ASAH) ASAH CHEM. IND CO LTD.
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO-NORDISK AS.
XX
XX Blinkovsky A, Brown K, Byun TS, Fujii M, Klotz AV;

PI	Marumoto C, Sloma A, Tang M;
XX	
DR	WPI: 1999-418769/35.
DR	P-PSDB: AAV22184.
XX	
PT	New isolated aminopeptidase polypeptides used in, e.g. food industry
XX	
PS	Claim 14; Fig 9; 8app; English.
XX	
CC	This sequence encodes the Sphingomonas capsulata IFO12533 aminopeptidase
CC	of the invention. The aminopeptidase polypeptides catalyse the removal of
CC	amino acids from the N-terminal end of peptides, oligopeptides or
CC	proteins. They can be used in the production of protein hydrolysates for
CC	enhancing the degree of hydrolysis and flavour development, particularly
CC	in foods. They can also be used to deactivate enzymes. They can also be
CC	used for specific cleavage of peptide sequences, e.g. to provide the
CC	necessary post-translational processing to activate precursor proteins.
XX	
SO	Sequence 3000 BP; 528 A; 978 G; 978 G; 516 T; 0 other;
	Query Match 0.6%; Score 18; DB 20; Length 3000;
	Best Local Similarity 100.0%; Pred. No. 1.4e+02;
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	2092 gctgtgtctgctggaagc 2109
DB	1911 gctgtgtctgctggaagc 1928
	RESULT 67
	AAH51776/C
ID	AAH51776 standard; DNA; 3001 BP.
XX	
AC	AAH51776;
XX	
DT	29-AUG-2001 (first entry)
XX	
DE	Chromosome 13q31-q33 biallelic marker containing amplicon SEQ ID 188.
XX	
XX	sbq1: g34665; sbq2: g35017; g35018; chromosome 13q31-q33; haplotype;
KW	biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200058510-A2.
XX	
PD	05-OCT-2000.
XX	
PF	30-MAR-2000; 2000WO-IB00435.
XX	
PR	30-MAR-1999; 99US-0126903.
PR	30-APR-1999; 99US-0131971.
PR	30-APR-1999; 99US-0132065.
PR	14-JUL-1999; 99US-0145928.
PR	27-JUL-1999; 99US-0145915.
PR	29-JUL-1999; 99US-0146452.
PR	29-JUL-1999; 99US-0146453.
PR	28-OCT-1999; 99US-0162288.
XX	
PA	(GEST) GENSET.
XX	
PI	Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Blhain B;
PI	Essioux L;
XX	
DR	WPI: 2000-619082/59.
XX	
PT	Polynucleotides comprising sequences from sbq1 and g35018 biallelic
PT	markers are used for genotyping and detecting schizophrenia or bipolar
XX	disorder and predisposition to these disorders -
PS	Claim 2; Page 680-681; 737pp; English.
XX	

CC	AAH51601	represents a human genomic nucleotide sequence comprising sbg1,
CC	g34665, sbg2, g35017 and g35018	nucleic acid sequences located on the
CC	human chromosome 13q31-q33	locus. The nucleotide sequences contain
CC	biallelic markers and polymorphisms.	Sequences AAH51602 - AAH51626 and
CC	AAH62907 - AAH62915	represent cDNA human sbg1 cDNA sequences and protein
CC	products. AAH51627 - AAH51631	and AAB62916 - AAB62918 represent g35018
CC	cDNA sequences and protein products.	Primers AAH51632 - AAH51699 are used
CC	to isolate sbg1 cDNAs, while sbg1	exons from different primates are
CC	represented by sequences AAH51642 -	AAH51699. Nucleotide sequences of
CC	amplicons which comprise biallelic	markers located on the chromosome
CC	13q31-q33 locus are represented in	AAH51700 -AAH51817. Biallelic markers
CC	are represented in the sequences by	degenerate/undefined base codes. PCR
CC	primers AAH51818 and AAH51819	are used in the isolation of sequences of
CC	the invention. The biallelic marker	containing nucleotide sequences are
CC	used to determine the identity of the	nucleotide at a biallelic marker in
CC	a sample DNA sequence. The nucleotide	sequences may be labelled and used
CC	for genotyping by determining the	identity of a nucleotide at a Region
CC	D-related biallelic marker in a	biological sample from single or multiple
CC	subjects. By determining the	frequency of a biallelic marker in a
CC	population an association between	a genotype and a trait, a haplotype and
CC	a trait and a phenotype and a	trait can be detected. The sequences can
CC	be used to determine a predisposition	to or early onset of schizophrenia or
CC	bipolar disorder or a beneficial	response to or side effects related to
CC	treatment against schizophrenia	or bipolar disorder.
CC	XX	
SQ	Sequence 3001 BP; 993 A; 463 C;	533 G; 1011 T; 1 other;
XX		
Query Match	0.6%;	Score 18; DB 21; Length 3001;
Best Local Similarity	100.0%;	Pred. No. 1.4e+02;
Matches 18;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
OY	1628 aagaatctgaaagcca	1645
DB	181 AAGAAATTCGTGAACCA	164
RESULT 68		
AAH51787/C		
ID	AAH51787	standard; DNA: 3001 BP.
XX		
AC	AAH51787;	
XX		
DT	29-AUG-2001	(first entry)
XX		
DE	Chromosome 13q31-q33	biallelic marker containing amplicon SEQ ID 199.
XX		
KW	sbg1; g34665; sbg2; g35017;	g35018; chromosome 13q31-q33; haplotype;
KW	biallelic marker; polymorphism;	schizophrenia; bipolar disorder; ds.
XX		
OS	Homo sapiens.	
XX		
PN	WO200058510-A2.	
XX		
PD	05-OCT-2000.	
XX		
PF	30-MAR-2000; 2000WO-IB00435.	
XX		
PR	30-MAR-1999; 99US-0126903.	
PR	30-APR-1999; 99US-0131971.	
PR	30-APR-1999; 99US-0132065.	
PR	14-JUL-1999; 99US-0143928.	
PR	27-JUL-1999; 99US-0145915.	
PR	29-JUL-1999; 99US-0146452.	
PR	29-JUL-1999; 99US-0146453.	
PR	28-OCT-1999; 99US-0162288.	
XX		
PA	(GSET) GENSET.	
XX		
PI	Cohen D, Blumenfeld M, Chumakov I,	Bougueleret L, Bihain B;
PI	Essioux L;	
XX		
DR	WPI: 2000-619082/59.	

XX Polynucleotides comprising sequences from sbp1 and g35018 biallelic
PT markers are used for genotyping and detecting schizophrenia or bipolar
PT disorder and predisposition to these disorders -
XX
PS Claim 2, Page 696-697; 737pp; English.

XX AAH51601 represents a human genomic nucleotide sequence comprising sbp1,
CC g3465, sbp2, g35017 and g35018 nucleic acid sequences located on the
CC human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC AAB62907 - AAB62915 represent cDNA human sbp1 cDNA sequences and protein
CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018
CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC to isolate sbp1 cDNAs, while sbp1 exons from different primates are
CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC amplicons which comprise biallelic markers located on the chromosome
CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers
CC are represented in the sequences by degenerate/undefined base codes. PCR
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC the invention. The biallelic marker containing nucleotide sequences are
CC used to determine the identity of the nucleotide at a biallelic marker in
CC a sample DNA sequence. The nucleotide sequences may be labelled and used
CC for genotyping by determining the identity of a nucleotide at a Region
CC D-related biallelic marker in a biological sample from single or multiple
CC subjects. By determining the frequency of a biallelic marker in a
CC population an association between a genotype and a trait, a haplotype and
CC a trait and a phenotype and a trait can be detected. The sequences can be
CC used to determine a predisposition to or early onset of schizophrenia or
CC bipolar disorder or a beneficial response to or side effects related to
CC treatment against schizophrenia or bipolar disorder.
XX
SQ Sequence 3001 BP; 996 A; 427 C; 613 G; 964 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 3001;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1628 aagaattctgaagcga 1645
|||||
Db 2866 AAGAAATTCGAAAGCCA 2849

RESULT 69
AAH25373
ID AAH25373 standard; DNA; 3275 BP.

XX AAH25373;

DT 15-MAY-2001 (first entry)

XX Genomic sequence of the Arabidopsis SGS3 gene.

KW SGS3 gene; post-transcriptional inactivation; RNA degradation;
KM viral resistance; resistance; fatty acid content; protein content; ss.
XX

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT exon 696..1658

FT intron /*tag= a 1659..1731

FT /*tag= b 1732..2023

FT exon /*tag= c 2024..2134

FT intron /*tag= d 2135..2379

FT exon /*tag= e 2380..2481

FT intron /*tag= f 2482..2648

FT exon

FT /*tag= g 2649..2738
FT intron /*tag= h 2739..2949
FT exon /*tag= i
XX
XX WO200105951-A2.

XX 25-JAN-2001.

XX 13-JUL-2000; 2000WO-FR02052.

XX 16-JUL-1999; 99FR-0009417.

XX 26-JAN-2000; 2000FR-0001006.

XX (AVERT) AVENTIS CROSCIENCE SA.

XX (INRG) INST. NAT. RECH. AGRONOMIQUE.

XX Beclin C, Elmayan T, Vaucheret H;

XX WPI: 2001-159529/16.

XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT resistance in plants and, when inhibited, for increasing transgene
PT expression -

XX Claim 1; Page 31-32; 36pp; French.

XX The present sequence represents the genomic sequence of the Arabidopsis
CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
CC inactivation (degradation of RNA) and for resistance to viruses.
CC Overexpression of SGS3 results in plants with increased resistance
CC to viruses, while inactivation of SGS3 in transgenic plants (e.g. by
CC expressing antisense RNA, by mutation or by homologous recombination)
CC increases the level of the transgene product. This product may e.g.
CC impart resistance (to herbicide, insects or pathogens), alter contents
CC of essential fatty acids or proteins, or is pharmaceutically active,
CC e.g. an immunoglobulin or interferon.

SQ Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 3275;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 985 aggtgcttgaggaatctc 1002
|||||
Db 2267 aggtgcttgaggaatctc 2284

RESULT 70

AAH31144
ID AAH31144 standard; cDNA; 3529 BP.

XX AAH31144;

DT 27-JUL-2001 (first entry)

XX Human colon cancer cell line Km12L4-A cDNA library derived seq#1078.

XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
KM detection; colon cancer cell line Km12L4-A; ss.

XX Homo sapiens.

XX WO200018916-A2.

XX 06-APR-2000.

XX 23-SEP-1999; 99WO-US222226.

XX 28-SEP-1998; 98US-0102161.

PR 28-SEP-1998; 9805-0102180.
 PR 29-SEP-1998; 9805-0102380.
 PR 08-OCT-1998; 9805-0103815.
 PR 27-OCT-1998; 9805-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Ctkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
 DR WPI; 2000-293155/25.
 XX
 PT Polynucleotide library comprising 1079 defined sequences, useful in
 PT the form of an array to detect cancer or susceptibility to cancer -
 PS Claim 1; Page 501-502; 502pp; English.
 XX
 PS The present invention describes a library of polynucleotides comprising
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
 CC are: (1) an isolated polynucleotide (1) having at least 90% identity to
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
 CC (3) an isolated polypeptide (11) encoded by (1); (4) an antibody that
 CC specifically binds to (11); (5) a vector comprising (1); and (6) a method
 CC of detecting differentially expressed genes correlated with a cancerous
 CC state of a mammalian cell comprising detecting a gene product encoded by
 CC 65 of the 1079 sequences given in the specification. The polynucleotides
 CC are used to monitor patients having (or susceptible) to cancer to detect
 CC potentially malignant events at a molecular level before they are
 CC detectable at a gross morphological level. The polynucleotides are also
 CC useful for monitoring the efficacy of various therapies and preventive
 CC interventions. Polynucleotide probes based on the disclosed sequences
 CC are useful for chromosome mapping and detection of transcription levels.
 CC The 1079 polynucleotide sequences were derived from a human colon cancer
 CC cell line Km124-A cDNA library.
 XX
 SQ Sequence 3529 BP; 1248 A; 778 C; 692 G; 803 T; 8 other;
 XX
 Query Match 0.6%; Score 18; DB 21; Length 3529;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2796 aaagaacctctgaaaaa 2813
 ||||||||||||||||
 Db 3465 aaagaacctctgaaaaa 3482
 XX
 RESULT 71
 AAV80604/c
 ID AAV80604 standard; cDNA; 3942 BP.
 XX
 AC AAV80604;
 XX
 DT 01-MAR-1999 (first entry)
 XX
 DE Kidney injury associated molecule HM050 cDNA clone.
 XX
 KW Kidney injury associated molecule; kidney injury related molecule;
 KW KIM; tissue growth promotion; regeneration; renal condition;
 KW acute renal failure; acute nephritis; tumour; ds.
 XX
 OS Rattus sp.
 XX
 PH Key Location/Qualifiers
 FT CDS 7..1329
 FT /*tag= a /product= "kidney injury associated molecule"
 FT /*tag= b /complement (3542..3733)
 FT /*tag= b /label= SAC_24644

XX
 XX WC09853071-A1.
 XX
 PN 26-NOV-1998.
 PD
 XX
 PF 22-MAY-1998; 98MO-US10547.
 XX
 PR 23-MAY-1997; 9705-0047491.
 PR 23-MAY-1997; 9705-0047490.
 XX
 PA (BIOJ) BIOGEN INC.
 XX
 PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
 PT WPI; 1999-045312/04.
 DR P-PDB; AAW86320.
 XX
 PT Kidney injury-associated molecule, KIM, polypeptides - upregulated
 PT in injured or regenerating tissues, useful to promote tissue growth
 PT and regeneration, especially to treat renal conditions
 PS Claim 9; Page 109-113; 213pp; English.
 XX
 PS The present sequence represents a kidney injury associated molecule
 CC (KIM) cDNA clone. KIM proteins can be administered therapeutically
 CC by expressing KIM encoding polynucleotides, to promote growth and/or
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
 CC are upregulated in injured or regenerating (especially renal) tissues.
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/dysregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal
 CC tissue growth arising from/affecting renal tissue). The proteins can
 CC also be used to locate KIM-producing cells (especially specific loci,
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
 CC arising from/affecting renal tissue), by contacting cells with an
 CC imageable KIM-binding reagent and imaging reagent accumulation.
 XX
 SQ Sequence 3942 BP; 908 A; 1019 C; 1087 G; 928 T; 0 other;
 XX
 Query Match 0.6%; Score 18; DB 20; Length 3942;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1883 cagaagacacagtgtaa 1900
 ||||||||||||||||
 Db 2495 CAGAAGACACAGGTGAA 2478
 XX
 RESULT 72
 AAV64473/c
 ID AAV64473 standard; DNA; 3969 BP.
 XX
 AC AAV64473;
 XX
 DT 02-JAN-2001 (first entry)
 XX
 DE DNA encoding an astaxanthin synthetase polypeptide.
 XX
 KW Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast;
 KW antioxidant; cancer; colouring reagent; farmed fish; salmon; ss.
 XX

Sequence 3969 BP; 937 A; 942 C; 869 G; 1221 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 3969;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 aacttgaaagacaccttc 402
|||||
DB 1193 AACCTGAAAGCACCTTC 1176

RESULT 73

AA158415
ID AA158415 standard; cDNA; 4215 BP.

AC AA158415;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 618.

Human; nootropic; immunosuppressant; cytosolic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukemia; ss.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;

PI: 2001-442253/47.

PI: P-SDB; AAM39259.

PT Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -

PT Claim 1; SEQ ID NO 618; 10078bp; English.

PS The invention relates to human nucleic acids (AA157798-AA161369) and
the encoded polypeptides (AAM38642-AA42213) with nootropic,

CC immunosuppressant and cytosolic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukemia and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
specification.

CC Sequence 4215 BP; 1523 A; 901 C; 855 G; 934 T; 2 other;

Query Match 0.6%; Score 18; DB 22; Length 4215;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2796 aaagaacctctgaaaaa 2813
|||||
DB 3248 aaagaacctctgaaaaa 3265

RESULT 74

AA160201
ID AA160201 standard; cDNA; 4871 BP.

AC AA160201;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 4190.

Human; nootropic; immunosuppressant; cytosolic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukemia; ss.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;

PI: 2001-442253/47.

PI: P-SDB; AAM41045.

PT Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -

PT Claim 1; SEQ ID NO 4190; 10078bp; English.

PS The invention relates to human nucleic acids (AA157798-AA161369) and
the encoded polypeptides (AAM38642-AA42213) with nootropic,

CC immunosuppressant and cytosolic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

SQ Sequence 4871 BP; 1815 A; 1021 C; 990 G; 1045 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 4871;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2796 aaagaacctcgtgaaaa 2813
 |||
 DB 3258 aaagaacctcgtgaaaa 3275

RESULT 75

AA13198/C
 ID AA13198 standard; DNA: 4951 BP.

AC AA13198;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:261.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.

PS Claim 1; Page 1248-1251; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AA12938 to AA13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.

SQ Sequence 4951 BP; 1563 A; 981 C; 735 G; 1665 T; 7 other;

Query Match 0.6%; Score 18; DB 20; Length 4951;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2492 gctgtcgtcaaatgcag 2509
 |||
 DB 1288 GCTGTCTGCANATGCAG 1271

Search completed: March 25, 2002, 16:52:45
 Job time: 9229 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Rdn on: March 25, 2002, 10:56:45 ; Search time 29.55 Seconds

(Without alignments)
2566.871 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 1024

Sequence: 1 MNFIKDSRAIQRMGMVTI.....MQFDDDLSTVITGAFKLVTA 1024

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

A.Geneseq_1101:*

1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	100.0	1024	22	AAU02880 Human caspase recr
2	837	80.8	1204	22	AAU02881 Human caspase recr
3	538	52.5	841	22	AAU02881 Human caspase recr
4	425	41.5	642	22	AAU02881 Human caspase recr
5	73	7.1	73	22	AAU02881 Human caspase recr
6	73	7.1	73	22	AAU02881 Human caspase recr
7	73	7.1	73	22	AAU02881 Human caspase recr
8	73	7.1	73	22	AAU02881 Human caspase recr
9	73	7.1	73	22	AAU02881 Human caspase recr
10	73	7.1	73	22	AAU02881 Human caspase recr
11	73	7.1	73	22	AAU02881 Human caspase recr

12	8	0.8	466	10	AAU02880 Human caspase recr
13	8	0.8	466	22	AAU02881 Human caspase recr
14	8	0.8	1112	17	AAU02881 Human caspase recr
15	8	0.8	1112	17	AAU02881 Human caspase recr
16	8	0.8	1112	17	AAU02881 Human caspase recr
17	8	0.8	1112	17	AAU02881 Human caspase recr
18	8	0.8	1112	17	AAU02881 Human caspase recr
19	8	0.8	1112	17	AAU02881 Human caspase recr
20	8	0.8	1112	17	AAU02881 Human caspase recr
21	8	0.8	1112	17	AAU02881 Human caspase recr
22	8	0.8	1112	17	AAU02881 Human caspase recr
23	8	0.8	1112	17	AAU02881 Human caspase recr
24	8	0.8	1112	17	AAU02881 Human caspase recr
25	8	0.8	1112	17	AAU02881 Human caspase recr
26	8	0.8	1112	17	AAU02881 Human caspase recr
27	8	0.8	1112	17	AAU02881 Human caspase recr
28	8	0.8	1112	17	AAU02881 Human caspase recr
29	8	0.8	1112	17	AAU02881 Human caspase recr
30	8	0.8	1112	17	AAU02881 Human caspase recr
31	8	0.8	1112	17	AAU02881 Human caspase recr
32	8	0.8	1112	17	AAU02881 Human caspase recr
33	8	0.8	1112	17	AAU02881 Human caspase recr
34	8	0.8	1112	17	AAU02881 Human caspase recr
35	8	0.8	1112	17	AAU02881 Human caspase recr
36	8	0.8	1112	17	AAU02881 Human caspase recr
37	8	0.8	1112	17	AAU02881 Human caspase recr
38	8	0.8	1112	17	AAU02881 Human caspase recr
39	8	0.8	1112	17	AAU02881 Human caspase recr
40	8	0.8	1112	17	AAU02881 Human caspase recr
41	8	0.8	1112	17	AAU02881 Human caspase recr
42	8	0.8	1112	17	AAU02881 Human caspase recr
43	8	0.8	1112	17	AAU02881 Human caspase recr
44	8	0.8	1112	17	AAU02881 Human caspase recr
45	8	0.8	1112	17	AAU02881 Human caspase recr
46	8	0.8	1112	17	AAU02881 Human caspase recr
47	8	0.8	1112	17	AAU02881 Human caspase recr
48	8	0.8	1112	17	AAU02881 Human caspase recr
49	8	0.8	1112	17	AAU02881 Human caspase recr
50	8	0.8	1112	17	AAU02881 Human caspase recr
51	8	0.8	1112	17	AAU02881 Human caspase recr
52	8	0.8	1112	17	AAU02881 Human caspase recr
53	8	0.8	1112	17	AAU02881 Human caspase recr
54	8	0.8	1112	17	AAU02881 Human caspase recr
55	8	0.8	1112	17	AAU02881 Human caspase recr
56	8	0.8	1112	17	AAU02881 Human caspase recr
57	8	0.8	1112	17	AAU02881 Human caspase recr
58	8	0.8	1112	17	AAU02881 Human caspase recr
59	8	0.8	1112	17	AAU02881 Human caspase recr
60	8	0.8	1112	17	AAU02881 Human caspase recr
61	8	0.8	1112	17	AAU02881 Human caspase recr
62	8	0.8	1112	17	AAU02881 Human caspase recr
63	8	0.8	1112	17	AAU02881 Human caspase recr
64	8	0.8	1112	17	AAU02881 Human caspase recr
65	8	0.8	1112	17	AAU02881 Human caspase recr
66	8	0.8	1112	17	AAU02881 Human caspase recr
67	8	0.8	1112	17	AAU02881 Human caspase recr
68	8	0.8	1112	17	AAU02881 Human caspase recr
69	8	0.8	1112	17	AAU02881 Human caspase recr
70	8	0.8	1112	17	AAU02881 Human caspase recr
71	8	0.8	1112	17	AAU02881 Human caspase recr
72	8	0.8	1112	17	AAU02881 Human caspase recr
73	8	0.8	1112	17	AAU02881 Human caspase recr
74	8	0.8	1112	17	AAU02881 Human caspase recr
75	8	0.8	1112	17	AAU02881 Human caspase recr
76	8	0.8	1112	17	AAU02881 Human caspase recr
77	8	0.8	1112	17	AAU02881 Human caspase recr
78	8	0.8	1112	17	AAU02881 Human caspase recr
79	8	0.8	1112	17	AAU02881 Human caspase recr
80	8	0.8	1112	17	AAU02881 Human caspase recr
81	8	0.8	1112	17	AAU02881 Human caspase recr
82	8	0.8	1112	17	AAU02881 Human caspase recr
83	8	0.8	1112	17	AAU02881 Human caspase recr
84	8	0.8	1112	17	AAU02881 Human caspase recr

Human muscarinic a	Cholinergic recept
Tomato pathogen re	Tomato pathogen re
Human ORFX ORF208	Human ORFX ORF208
Human bone marrow	Human bone marrow
Human INTERCEPT 29	Human INTERCEPT 29
Human integrin bet	Human integrin bet
Peptide #6693 enco	Peptide #6693 enco
Peptide #7244 enco	Peptide #7244 enco
Peptide #10278 enc	Peptide #10278 enc
Streptococcus pneu	Streptococcus pneu
Amino acid sequenc	Amino acid sequenc
Human prostate can	Human prostate can
Human secreted pro	Human secreted pro
Branched chain 2-o	Branched chain 2-o
Human prostate can	Human prostate can
Amino acid sequenc	Amino acid sequenc
Arabidopsis thalia	Arabidopsis thalia
Human -212 cadherin	Human -212 cadherin
Protocadherin clon	Protocadherin clon
Pinus radiata cell	Pinus radiata cell
Human steroid horm	Human steroid horm
Arabidopsis thalia	Arabidopsis thalia
Vancomycin resista	Vancomycin resista
Arabidopsis thalia	Arabidopsis thalia
Group B Streptococ	Group B Streptococ
Human olfactory re	Human olfactory re
Corynebacterium gl	Corynebacterium gl
Arabidopsis thalia	Arabidopsis thalia
Arabidopsis thalia	Arabidopsis thalia
Arabidopsis thalia	Arabidopsis thalia
S. pneumoniae cell	S. pneumoniae cell
Arabidopsis thalia	Arabidopsis thalia
Human secreted pro	Human secreted pro
Human hyaluronase-11	Human hyaluronase-11
Human ORFX ORF2943	Human ORFX ORF2943
Human Electron Tra	Human Electron Tra
P501S-specific T c	P501S-specific T c
C glutamicum prote	C glutamicum prote
Human high voltage	Human high voltage
Voltage-dependent	Voltage-dependent
Protein encoded by	Protein encoded by
Human OR-like poly	Human OR-like poly
Arabidopsis thalia	Arabidopsis thalia
Arabidopsis thalia	Arabidopsis thalia
Human G protein co	Human G protein co
Streptococcus pneu	Streptococcus pneu
Arabidopsis thalia	Arabidopsis thalia
Arabidopsis thalia	Arabidopsis thalia
Amino acid sequenc	Amino acid sequenc
Actinoplanes sp. a	Actinoplanes sp. a
Fragment of human	Fragment of human
Arabidopsis thalia	Arabidopsis thalia
Escherichia coli R	Escherichia coli R
Human INTERCEPT 29	Human INTERCEPT 29
Campylobacter jejuni	Campylobacter jejuni
PDC-E2, BCOADC-E2,	PDC-E2, BCOADC-E2,
Acyl-coenzyme A:ch	Acyl-coenzyme A:ch

85	7	0.7	370	22	AAU00243	Human histone deac
86	7	0.7	371	22	AAG81767	S. epidemidis ope
87	7	0.7	374	22	AAG11348	Arabidopsis thalia
88	7	0.7	382	22	AAG91112	C glutamicum prote
89	7	0.7	391	21	AAG11347	Arabidopsis thalia
90	7	0.7	405	22	AAB94128	Human protein sequ
91	7	0.7	411	22	AAG92713	C glutamicum prote
92	7	0.7	415	21	AAB10998	M. crenulata hemoc
93	7	0.7	419	22	AAB71110	M. crenulata KLH1
94	7	0.7	420	22	AAB71153	M. crenulata KLH1
95	7	0.7	425	22	AAG82755	S. epidemidis ope
96	7	0.7	433	19	AAW76411	Human betac cytopl
97	7	0.7	437	20	AAV34786	Amino acid sequenc
98	7	0.7	439	22	AAG92885	C glutamicum prote
99	7	0.7	442	22	AAB83331	Corn phenylalanyl-
100	7	0.7	449	21	AAB36885	S. pneumoniae Yycg

ALIGNMENTS

RESULT 1

AAU02880 standard; Protein: 1024 AA.

AAU02880:

12-SEP-2001 (first entry)

Human caspase recruitment domain 12 (CARD-12) polypeptide #1.

Caspase recruitment domain: CARD-12; apoptosis; stress-related pathway;
cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
systemic lupus erythematosus; arthritis; neurological disorder; stroke;
Alzheimer's disease; amyotrophic lateral sclerosis; haematologic disease;
aplastic anaemia; myocardial infarction; inflammatory disorder;
Crohn's disease; insulin-dependent diabetes; contact dermatitis;
psoriasis; graft rejection; bacterial infection; leptomalous leprosy;
tuberculosis; ischaemic brain injury; hypoxic brain injury;
kidney ischaemia; reperfusion injury; acute bacterial meningitis;
excitotoxic brain damage; liver disease.

Homo sapiens.

MO200130971-A2.

03-MAY-2001.

26-OCT-2000; 2000WO-US29643.

27-OCT-1999; 99US-0161822.

(MILL-) MILLENNIUM PHARM INC.

Bertin J, Robison KE;

WPI: 2001-308628/32.

N-PSDB: AAS03945.

Isolated caspase recruitment domain-12 polypeptide and nucleic acids
encoding them, useful for treating and diagnosing disorders associated
with abnormal apoptosis such as cancer, arthritis and Alzheimer's
disease -

Cclaim 9; Fig 1; 93pp; English.

The sequence represents a human caspase recruitment domain 12 (CARD-12)
polypeptide. CARD domains are found in a number of proteins that transmit
signals that activate apoptosis and inflammatory pathways in response to
stress and other stimuli. Therefore, CARD-12 and its corresponding
c nucleic acid may be used in treatment and diagnosis of patients suffering
from disorders associated with an abnormal level (an increase or a
decrease) of apoptotic cell death or abnormal activity of stress-related

CC	pathways. The disorders include cancer, viral infections (e.g. caused by
CC	poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus
CC	erythematosus, arthritis), neurological disorders (e.g. Alzheimer's
CC	disease, amyotrophic lateral sclerosis), haematologic diseases (e.g.
CC	aplastic anaemia, myocardial infarction, stroke), inflammatory and immune
CC	system disorders (e.g. Crohn's disease, insulin-dependent diabetes,
CC	contact dermatitis, psoriasis, graft rejection), bacterial infections
CC	(e.g. tuberculosis, leptomalous leprosy), ischaemic and hypoxic brain
CC	injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage,
CC	acute bacterial meningitis and liver disease.
XX	
SQ	Sequence 1024 AA;
Query Match	100.0%; Score 1024; DB 22; Length 1024;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1024;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MNFIDNSRALLQRMGAYVYKQITDLEFVWVNLNREVNITICCEVEQDANGIIMLIK 60
DB	1 mnlfdnsrallqrmgtvkvkqtdldlfvwnvlnreevnlcekvqdaargllhmlik 60
QY	61 KGESCNLFKSLKEMVNPFLFODLNGOSLFHOTSGDDDLAODLKDIYHPPSFINFVPL 120
DB	61 kgescnlfkslkemvnpflfqdlngsflfqsegslddlagdkldlyhpsflnfypl 120
QY	121 GEDIDIFLNKSTPEVPLMRKQHHRRVEQLTLNGLQALQSPICIIIGESGKSTLLO 180
DB	121 gedidiflnkstfepylwrkdqghrveqltlnqllqalspsicigesgkystllq 180
QY	181 RIAMMGSKCALRKRFVFFRLSRAGGLFELFCQDLDPGTIRKQTFMAMLLKR 240
DB	181 riammgskcalrkrfvflrlsraggllfelcdqldipgtirkqtmanllkr 240
QY	241 QVFLFLQGVNFRPQNCPEIFALKENHREKNWYVTTTTECLHIFOGALTRAEVDM 300
DB	241 qvflflqgvnfrpncpeifalkenhreknwvtttclhifogaltaevdm 300
QY	301 TEDSAQALIREVYLKELAGLLQIQKSRCLNMLKTRPLFVVITCAIQMGSEFSHTQT 360
DB	301 tedsaqalirevylkelaglllqkscrlnmlktrplfvvitcaiqmgsefshqgt 360
QY	361 TLFHFTFDLLTQKNKHKKRGAASDFIRSLDHCGLAIEGVSRKFPDELQDVSSVNDY 420
DB	361 tlfhftfdlllqknkhkgyvaasdfirslhcgdlaiegvfshkfdelqdvssvndv 420
QY	421 LTTGLLCKYTPAQRPKPKFFKSFQETAGRRLSLTSHPEPVKNGVYLOKMYSI 480
DB	421 lltgllckytaqrkpkpkykffhksfgeytagrrlssltshepevltkngyiqkmysi 480
QY	481 SDITSTYSSLLRKYTCGSSVEATRAVWKHLAAYVQHCGLLSIARKPLMROESLOSXNT 540
DB	481 sditstysllrkytcgssveatravwmkhlaayvgcglglsiakrplmrweslosxnt 540
QY	541 TEOELTKAININSFEVCGHILYQESTSKSALSQEFARFQGSILYNSGNIPDYLFDFE 600
DB	541 teogeltkaininsfvegihlyqestsksalsqefarfgksilynsgnipdyldfde 600
QY	601 HLPNCASALDFIKLDFYCGAMASWEKAEDTGCINMEAPERYISRAVSLFFNNKQEFR 660
DB	601 hlpncasaldfikldfygamaswekaeedvglimeapeyisravsifnnkqefr 660
QY	661 TLEVLTRDFSKLNKODITYLGRIFSSATSLRLQIRKAGVAGSLVLSTCKNIYSLWE 720
DB	661 tlevltrdfsklnkodylgrifssatslrlqirkagvagslavlstckniyslwe 720
QY	721 ASPLTIEDEBRTSTYNTKITSIHDLQORLPQGTLSLGNKNTKILMNIKNEDDA 780
DB	721 aspltiebdrhtstynktisihdlqorlpqgtlsldgnkntkilmniknedda 780
QY	781 IKLAGSLKNKKMCLFHTLTHSDIGEGMDYIVKSLSSPCDLFEQIYVSSCLSANAVYL 840
DB	781 iklaegslknkkmcflthlthsdigegmdyivkslsspcdlfeiqvssclsanavyl 840

QY	841	AQNLENHVLKSLTIDASENTLEKDGNAALHELDIRMANVLEDTALMLPWCDDVGSLSLL	900
QY	841	agqnlhnlvklslsldisensylekdgnealheildirmvleqtlalmjpwgdcvqgsllsll	900
QY	901	KHLEEVPOLVKIGLKMWRRLTDETRILGAFEGKNPLKNFOOLNAGRRVSSDGLAAMGV	960
QY	901	khlveevpqvlvkgllkmwrltdeirrlilgafegknplknfqqnlhagrrvssdgvlamgv	960
QY	961	FENLKLQVFEFDESTKEFELDPALVRKLSQVLSKLTFLEQEARLVGWCFDDDLVSITGAFK	1020
QY	961	fenlklqvfeffdstkeffelpdpalvrklsqvlslkltfgearlvgywcfdddlsvitgafk	1020
QY	1021	LVYA 1024	
QY	1021	lvta 1024	
Db	1021	lvta 1024	
RESULT 2			
AAU02881			
ID	AAU02881	standard; Protein; 1204 AA.	
XX	AAU02881;		
AC			
XX			
DT	12-SEP-2001	(first entry)	
DE			
XX		Human caspase recruitment domain 12 (CARD-12) polypeptide #2.	
KW		Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;	
KW		cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;	
KW		systemic lupus erythematosus; arthritis; neurological disorder; stroke;	
KW		Alzheimer's disease; amyotrophic lateral sclerosis; haematologic disease;	
KW		aplastic anaemia; myocardiatal infarction; inflammatory disorder;	
KW		Crohn's disease; insulin-dependent diabetes; contact dermatitis;	
KW		psoriasis; graft rejection; bacterial infection; lepromatous leprosy;	
KW		tuberculosis; ischaemic brain injury; hypoxic brain injury;	
KW		kidney ischaemia; reperfusion injury; acute bacterial meningitis;	
KW		excitotoxic brain damage; liver disease.	
XX			
OS		Homo sapiens.	
PN		WO200130971-A2.	
PD		03-MAY-2001.	
XX		26-OCT-2000; 2000MO-US29643.	
XX		27-OCT-1999; 99US-0161822.	
PA		(MILL-) MILLENNIUM PHARM INC.	
PI		Bertin J, Robison KE;	
DR		WPI; 2001-308628/32.	
DR		N-PSDB; AAS033946.	
PT		Isolated caspase recruitment domain-12 polypeptide and nucleic acids	
PT		encoding them, useful for treating and diagnosing disorders associated	
PT		with abnormal apoptosis such as cancer, arthritis and Alzheimer's	
PT		disease -	
PS		Disclosure; Fig 2; 93pp; English.	
XX			
CC		The sequence represents a human caspase recruitment domain 12 (CARD-12)	
CC		polypeptide. CARD domains are found in a number of proteins that transmit	
CC		signals that activate apoptosis and inflammatory pathways in response to	
CC		stress and other stimuli. Therefore, CARD-12 and its corresponding	
CC		nucleic acid may be used in treatment and diagnosis of patients suffering	
CC		from disorders associated with an abnormal level (an increase or a	
CC		decrease) of apoptotic cell death or abnormal activity of stress-related	
CC		pathways. The disorders include cancer, viral infections (e.g. caused by	
CC		poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus	
CC		erythematosus, arthritis), neurological disorders (e.g. Alzheimer's	

Query Match	80.8%	Score 827	DB 22	Length 1204
Best Local Similarity	99.9%	Pred. No. 0		
Matches 927	Conservative	0	Mismatches 1	Indels 0
				Gaps 0
89	LEHQRSEGDLDADLADKLYTPSPFLNTPYRPGEDIDIDFNLSKSTPEEVLNMRKDQHNHR	148		
246	lfhqsegddlddaqlkdlhyhtpflnfpilrpygedididfnlksfdeplyrwdqnhhr	305		
149	VEQTLTLNGLAQLSPCIIEGESGKSTLLORIAMLMWSGSKALTKRPFVEFLRSRA	208		
306	veqtltnlglqlspcillegsqgkstllqriamlwsgskaltkfkfveflrsra	365		
209	QGGLEFETLCDQLDPGTRTKOTFMAMLLKQRYVFLFDGNERKPOUCPELEALIKEN	268		
366	qgglfetlcdqlldpgrtkqtfmamlklyrvyflldgynefkxpcpelaaliken	425		
269	HEFKMNVITYTTEECIRHTRQFGALTAEGYDMTDEBSAOLIREVYLKELAEGLLIQKS	328		
426	hfkfmnvitytteectirhqfaltaeaygdmtdesaqalirevylkelaegllylqkxs	485		
329	RCRLNMLKPTLFVTVTCALQMGSESEFHSHTQTLTFEPFDLLIQKNKHKHKAASDFLR	388		
466	rcrlnmlkptlfvtvtcaltqmgesehshqtqtlftfydlliqrknhkbgvaasdfir	545		
388	SLDHGCDLALBSVFESHKDFELQDYVSVNEDVLLTTGLCKTKTAORFKRYKFEHKSFOE	448		
546	slhdgcdalalegvfshkdfelqdyvsvnedvllttgllckxyaqfkipkykfhsqfe	605		
449	YVAGRRLSLSLSHPEEYTKNGYLOKQWVSLSDITSYVSLILRYTQGSVEATRVVMKH	508		
606	yvaqrslslshpeeytkngyloqkwvslsditsyvsllrytqgsveatrvvmkh	665		
509	IAAVYOHGCGLGSLIAKRPYLMROESLQSVKNTDEDEILKAININSEVEGCIHLYOESTSK	568		
666	laavyohgclglislakrplylwrqeslqsvkntdegeilkaaininstvegcilhlygestsk	725		
568	SALSGEPEAFQOGKSLYINSQNIPOYLEDPEHLPNCASALDPFKIDPYGGAWSAEKKA	628		
726	salsgepeafqgkslyinsgnipdyldfpehlpncasaldpfikidpyggamaswekka	785		
629	EDTGCIRHMEAEAEYTPRAVSLFPMWQKEPFTLEVTLPDESKLNQODITYYLGITSSAT	688		
786	edtgcirhmeaeaeyptraavslfpmwqkeftlevtlpdsklnqodityylygkltssat	845		
688	SLRLQIKRCAGVAGSLSLVSTCKNYSILMVEASPLTIEDERHITSVNMILKTLISIHDLON	748		
846	slrlqikrcagvagslslvstcknysilmveaspltiiederhitsvnmilktlslshdlqn	905		
749	ORLPGLTDSIGNLKNLTKYLMNDIKMNEDEAIKTAEGIKNLKMKCLFHLTHLSDIGEGM	808		
906	orlpgltdsiglnlknltkylmndikmneedaiktaegiknlkkmclfhlthlsdigegm	965		
809	OYIVYSLSSEPPDDEKEIOYVSCCLSNANVKTIAQONHNLYKISIIIDLSENYLEKQNEAL	868		
966	oyivyslsseppddeeioyvscclsnanvktiaqonhnlvklisliidlsenylekqneal	1025		
868	HELIIRMNVLQDLTALMLPWCDDVOGSSLSLKLHEEYQOLVYLGKAMNRLTDETRIRLG	928		
1026	heliirmnvlqdltalmlpwcddvogsslsllkhlheeyqolvylgkkmnrltdeetrirlg	1085		
929	AEFGKNPLKNFOOLNLAGNRVSSDCWLAEMGVFENLKOLVFFDFSTKEFLPDPALVRKLS	988		

Db 1086 affgknplknfqqldnagrnvasdglafmgvfenkqlvffdfsrckeflpdpalvrkis 1145
 QY 989 QVLSKTLFLOEARLYGWOPDDDDLSVIT 1016
 Db 1146 qvlskltflgearlvqwgfdiddldsvlt 1173

RESULT 3
 AAM23595
 ID AAM23595 standard; Protein: 841 AA.
 XX
 AC AAM23595;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Murine EST encoded protein SEQ ID NO: 1120.
 XX
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition.
 XX
 OS Mus musculus.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PE 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 XX
 PR 17-JUL-2000; 2000US-0617746.
 PR 13-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Dermanac RA, Zhang J, Werhman T;
 XX
 DR WPI: 2001-476164/51.
 DR N-PSDB: AAH98254.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 20; Page 832-833; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 XX
 SO Sequence 841 AA;

Query Match 52.5%; Score 538; DB 22; Length 841;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 838; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 184 MWMGSGKALTRKFEVFLRLSRAQGLFETLCDQLDIPGIRKQTFMMLTKLRQV 243
 Db 1 mltwsgkcaltkfkfvlflrlsraqglfctcdqllidipgtrtkqtfmmltklrqv 60

QY 244 LFLDGVNEFRPQNCPEIEALIKENHRKNWYVTTTTECLRHTRQFQALRAVGDWMT 303
 Db 61 yflldgynefrpncpeieaalkenhrkwnwyvtttctclrhtrqfqaalraevgdwmed 120

QY 304 SAQALIREVLIKEALGELLGIQSRCLRNLMKRPFLVIVICAIQMGSEFHSHTQTLLF 363

Db 121 saqallirevlikelaegllilqiksrclrnlnktpflvvlcalqmgeseffshqtllf 180
 QY 364 HTEFYDILLIOKNKHKHKGVAASDFIRSLDHCGDLALGVFSHKFDFLODVSSVNEEDVLT 423
 Db 181 hctfydillqknkhkhkgvaasdfirslthcgyalalegvfshkfdefqdvssvneevllt 240

QY 424 TGLLCKYTAQRFKPKYFFHKSFQEYTAGRRLSSLTSHPEPVTKGNVLOKWSISDI 483
 Db 241 tgllickytagrfkpkkyffhkafqeytagrlssltshepevltkngyltkmwsisdi 300

QY 484 TSTYSSLLEVTGCGSSVEARRAVWKHLAAVYQGCCLGLSIAKRPIMKROSLDSVKNITTEQ 543
 Db 301 tstyssslrvtcgssveaeravwkhlavayqgcclglslakrplwrqeslvknitlq 360

QY 544 EILKAININSFVCEGCIHLQESTSKALSQEFEPFQCKSLTINSNDIVLFDPEHP 603
 Db 361 eilkaainnsfvecgihlyqestsksalsqefefqgkslyinsndivlfdfehp 420

QY 604 NCASALDFIKLDFYGGAMASWEKAEDTGTGIMEAPETVIPSRAVSLFPMNKQEFFRLE 663
 Db 421 ncasaldfiklgyggamaswekaeedtggtgimeapevtyipsravslfpmnkqeffrle 480

QY 664 VTLRDESKLNKODITYLCKIFSSATSRLQIKRCAGVASLSLVSTCKNITSIMVLEASP 723
 Db 481 vtlrdesklmkqdlrlylglkifssatsrlqkrcaqvagslslvstcknyslmveasp 540

QY 724 LTFEDERHITSVTNLTSTISHDLQNRPLGCLTDSIGNKNTLKLIMDKNEEDATKL 783
 Db 541 ltfeederhltstvtnlstishdlngprlpbgldtsignkntklkimdkneedaikl 600

QY 784 AEGKLNKKMCLFHLTHSDIGEMDYIVKSLSEPCDLEETQVSCISANAVKILAQN 843
 Db 601 aegklnkkmcclfhltshsdigemdyivksissepcdleeqivscisaavkilaqn 660

QY 844 LHNLVKLSITDISENVLEEDGNEALHETIDRNNVLEQLTALMLPMGCVOGSLSELKHL 903
 Db 661 lhnlvklsitdisenvleedgnealhelidrmnvleqltalmlpmgcvgqslsikhhl 720

QY 904 EEVPOLVRLKGLKNRRLTDEIRIILGAFEGKNPLKNFQOLNLAGNVSSDGMVLAFGVEEN 963
 Db 721 eevpolvrlkglknwrltdeiriligafegknpkfnqglndnlagrnvasdglafmgvfen 780

QY 964 LKQVFPDEFSTKEFLPDPALVAKLSQVLSKTLFLOEARLYGWOPDDDDLSVITGAFKLV 1023
 Db 781 lkqvlfpdefstkeflpdpalvrklsqvlskltflgearlvqwgfdiddldsvltgafklyt 840

QY 1024 A 1024
 Db 841 a 841

RESULT 4
 AAM25640
 ID AAM25640 standard; Protein: 642 AA.
 XX
 AC AAM25640;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:1155.
 XX
 KW Human; cancer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; vituicide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiagregant; haemostatic; vulnerrary; antilucer; osteopathic; eczema;
 KW dermatological; antiallergic; antiaslthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anapylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;

KW	thrombocytopopenia; osteoporosis; severe combined immunodeficiency;
KW	allergic rhinitis; diabetes; multiple sclerosis; depression;
KW	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder
KW	neurological disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200153455-A2.
XX	
PD	26-JUL-2001.
XX	
PF	22-DEC-2000; 2000WO-US35017.
XX	
PR	23-DEC-1999; 99US-0471275.
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT;
XX	
DR	WPI; 2001-457603/49.
DR	N-PSDB; AAH99581.
XX	
PT	Isolated human polynucleotides encoding polypeptides, useful for the
XX	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX	Claim 20; Page 240; 1217pp; English.

CC AAH919166 to AAH9904 encode the human proteins given in AAM25225 to
CC AAM25653. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antitumagen;
CC cardiovascular; antihaemic; antigenregant; haemostatic; vulnerary;
CC anticancer; osteoplastic; dermatologic; antiallergic; antiaesthetic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmune, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

SQ Sequence 642 AA;

Query Match	41.5%	Score 425;	DB 22;	Length 642;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 625; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 398 LEEFSEHKFDELODVSSVNEVDLLTTLGLCKTQAQFRRKRYKFEHKFSEQFYAGRLSS 457

Db 16 lEgFsmhfdIdelqdvssvmedvlltqlclckYaqgfkpkYkfifkbsIdEYlaqrllss 75

QY 458 LLTSHPEEYKNGCYLQKMWISIDITSPYSSLLRYCQSSSVATERAMKHLAAVYOHGC 517

Db 76 lltshpeeyvkngyYlqkmsIsIdItSYssllrycgsSvEaTraVmKhlaayYhgc 135

QY 518 LGLSLAKRPLMROESLQSYKNTTEQELKAININSEFECGHLHYOESTSKALSQEFEA 577

Db 136 llgIsIakRplmRwEsgIsqYvnttEgElKaInInSEfEcGhLhYoeStSkAlSqeFea 195

QY 578 FFGQKSLYINSQNTPOVLPDFEHLPRCAsALDQFIKIDFEGGAMASWEKRAEDTGGIHME 637

Db 196 ffgqkslylnsqnTpdvlpdfIEhlprCAsaldfIkIgYggamaSwEkRaEdTggIHme 255

Qy	638	EAERTYIPSRASVLEFNNMKOEFRLEVTLPBFSKLNKQDITYGKTRSSATSLRLQKRC	697
Db	256	eaepelyiprsavslifmwkgelrlevltdifsklnmqditylqkissatsrlqlkrc	315
Qy	698	AGVAGSLVSTWCKNIIYSLWVEASPLTIEDERHITSVTNLKFTSIHDLNORPGCLTD	757
Db	316	agvagslsvstwcknlyslwveaspltiiederhitsuclktsihdlngqrlpgcltd	375
Qy	758	SLGNLKNLTJLIMDNIKKNBEDATIKLBEGLKNLKKMLFHLTHLSDIGCEMDYIVYSSLS	817
Db	376	slgnlknltjlmndikmneedaiklaeglnlkkmlfhlthlsdigemdyivysls	435
Qy	818	EPDDEIEIQLYSCLSANAYVILNQNHLNLYKSLIDLSENYLEKQENLHLELIDRMY	877
Db	436	epcddeeiqlyscclsanavvllhqnlnlyklsldlsenylekdneahelldrmny	495
Qy	878	LEOLTALMLPWGCVOGSSLSLKHLEBPOLVLTGLKNMRITDTERITIGAFEGKNPK	937
Db	496	leqltalmlpwgcvgvgsslsllhleeppolvltglknmrldteritigafegknplk	555
Qy	938	NFOQLINLAGNFRSSDGMWLEFMGVFENLKOLVFPDPSRKEFLPDPALVRLSQUALSKTEL	997
Db	556	nfgqlinlagnrssdgmwlefmgvfenlkqlvfpdpsrkeflppdalyrklsgualsktfl	615
Qy	998	QEARLVGMQPDDBDLVYTGAFAKLVTA	1024
Db	616	qearlvqwgfdaddslvltgaflklyta	642

RESULT	5
AAM17792	
ID	AAM17792 standard; Protein; 73 AA

DT 12-OCT-2001 (first entry)

DE Peptide #4226 encoded by probe for measuring cervical gene expression.

KM Probe; human; microarray; gene expression; cervical epithelial cell; cervical cancer.

OS	Homo sapiens.
XX	
PN	W0200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-01080312
PR 26-MAY-2000; 2000US-0207456
PR 30-JUN-2000; 2000US-0608408
PR 03-AUG-2000; 2000US-0632366
PR 21-SEP-2000; 2000US-0234687
PR 27-SEP-2000; 2000US-0236359
PR 04-OCT-2000; 2000GB-0024263

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR

DR WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for

PS Claim 27; SEQ ID No 22618; 487pp; English.

CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see A110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs

CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
SQ Sequence 73 AA;

Query Match 7.1%; Score 73; DB 22; Length 73;
Best Local Similarity 100.0%; Pred. No. 6.9e-66;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 944 LAGNRVSSDGLAFMGVFNELKQVFEFDESTEKFLPDPALVKKLSQVLSKTLFLQEARLV 1003
DB 1 lagnrsvssdgwlafrngvfenlkqvlfdftskelflppdpalvkrklsqvskskltflgearlv 60
OY 1004 GWOFFDDDLSTVIT 1016
DB 61 gwqfdddlstvit 73

RESULT 6
AAM30302
ID AAM30302 standard; Protein; 73 AA.
XX
AC AAM30302;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #4339 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KM genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PS (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 30571; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs;
CC see AAI3115-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 73 AA;

Query Match 7.1%; Score 73; DB 22; Length 73;
Best Local Similarity 100.0%; Pred. No. 6.9e-66;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 944 LAGNRVSSDGLAFMGVFNELKQVFEFDESTEKFLPDPALVKKLSQVLSKTLFLQEARLV 1003
DB 1 lagnrsvssdgwlafrngvfenlkqvlfdftskelflppdpalvkrklsqvskskltflgearlv 60
OY 1004 GWOFFDDDLSTVIT 1016
DB 61 gwqfdddlstvit 73

RESULT 7
AAM05439
ID AAM05439 standard; Protein; 73 AA.
XX
AC AAM05439;
XX
DT 09-OCT-2001 (first entry)
XX
DE Peptide #4121 encoded by probe for measuring breast gene expression.
XX
KW Probe; human; breast disease; breast cancer; development disorder;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PS (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 27; SEQ ID No 14179; 322bp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes
CC (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 73 AA;

Query Match 7.1%; Score 73; DB 22; Length 73;

Best Local Similarity 100.0%; Pred. No. 6,9e-66;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 LAGNRSSDGLFMGFEELKQLVFFDFSTKEFLPDALVRKISQVLSKLTFLQEARLV 1003
Db 1 lagnrssdglfmgyfienlkqivfdistkeflpdalvrklsqylskltflqearlv 60

QY 1004 GWOFFDDDLVSIT 1016
Db 61 gwqfdddlsvit 73

RESULT 8
AAW98555
ID AAW98555 standard; Protein: 1021 AA.
XX AAW98555;
AC AAW98555;
XX
XX 31-MAR-1999 (first entry)
XX
XX H. pylori GHPO 1661 protein.
DE
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
XX Helicobacter pylori.
OS
XX W09843478-A1.
PN
XX 08-OCT-1998.
PD
XX 01-APR-1998; 98WO-US06371.
PF
XX 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX Al-Garawi A, Kleantchous H, Miller C, Oomen RP, Tomb J;
PI WPI; 1998-542293/46.
DR N-PSDB; AAX14274.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
XX
XX Claim 8; Page 1082-1087; 2054pp; English.
PS
XX This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
XX Sequence 1021 AA;
SQ

Query Match 0.9%; Score 9; DB 19; Length 1021;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 735 VTNLKTLSI 743
Db 879 vtnlktlsi 887

RESULT 9
AAV02693

ID AAV02693 standard; Protein: 110 AA.
XX AAV02693;
AC AAV02693;
XX
XX 11-JUN-1999 (first entry)
DT
XX Human secreted protein encoded by gene 44 clone HTDAD22.
XX
DE
XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; flissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
XX Homo sapiens.
OS
XX W09902546-A1.
PN
XX 21-JAN-1999.
PD
XX 07-JUL-1998; 98WO-US13684.
PF
XX
XX 12-SEP-1997; 97US-0058785.
PR 08-JUL-1997; 97US-0051916.
PR 08-JUL-1997; 97US-0051818.
PR 08-JUL-1997; 97US-0051919.
PR 08-JUL-1997; 97US-0051920.
PR 08-JUL-1997; 97US-0051925.
PR 08-JUL-1997; 97US-0051926.
PR 08-JUL-1997; 97US-0051928.
PR 08-JUL-1997; 97US-0051829.
PR 08-JUL-1997; 97US-0051930.
PR 08-JUL-1997; 97US-0051931.
PR 08-JUL-1997; 97US-0051932.
PR 08-JUL-1997; 97US-0052732.
PR 08-JUL-1997; 97US-0052733.
PR 08-JUL-1997; 97US-0052793.
PR 08-JUL-1997; 97US-0052795.
PR 08-JUL-1997; 97US-0052803.
PR 18-AUG-1997; 97US-0055684.
PR 18-AUG-1997; 97US-0055722.
PR 18-AUG-1997; 97US-0055723.
PR 18-AUG-1997; 97US-0055947.
PR 18-AUG-1997; 97US-0055948.
PR 18-AUG-1997; 97US-0055949.
PR 18-AUG-1997; 97US-0055950.
PR 18-AUG-1997; 97US-0055953.
PR 18-AUG-1997; 97US-0055954.
PR 18-AUG-1997; 97US-0055964.
PR 18-AUG-1997; 97US-0055984.
PR 18-AUG-1997; 97US-0056360.
PR 12-SEP-1997; 97US-0058660.
PR 12-SEP-1997; 97US-0058661.
PR 12-SEP-1997; 97US-0058664.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA;
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;
XX WPI; 1999-120770/10.
DR N-PSDB; AAX27354.
XX
XX New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 11; Page 343; 464pp; English.
PS
XX This sequence represents a secreted human protein encoded by the gene
CC

CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AAX27302) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 123 novel genes and their fragments (nucleic
CC acid sequences: AAX27311-X27449; amino acid sequences AAX02650-102788)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 123
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX27311 for described uses).
XX
SQ Sequence 110 AA;

Query Match 0.8%; Score 8; DB 20; Length 110;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 FVFPLRLS 206
| | | | | | | |
Db 21 fvfflrlls 28

RESULT 10
AAR48689
ID AAR48689 standard; Protein; 317 AA.
XX
AC AAR48689;
XX
DT 05-JUN-1996 (first entry)
XX
DE G-protein coupled human m2 muscarinic acetylcholine receptor protein.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW psychotic disorder; schizophrenia; dopamine; CAMP; adenosine; thrombin;
KW muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW rhodopsin; opsin; odorant; cytomegalovirus.
XX
OS Homo sapiens.
XX
PN W09405695-A1.
XX
PD 17-MAR-1994.
XX
PF 09-SEP-1993; 93WO-US08528.
XX
PR 10-SEP-1992; 92US-0943236.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Murphy RB, Schuster DI;
XX
DR WPI: 1994-101120/12.
XX
PT Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PT binding GPR ligands or modulating GPR binding
XX
PS Disclosure: Page 61; 160pp; English.
XX
SQ Proteins AAR48685-R48758 represent a range of G-protein coupled receptor
CC proteins selected from CAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomegaloviral and other G-protein coupled receptors. The
CC receptor proteins were used to design polypeptides, pref. based on the
CC transmembrane domains, for use in G-protein coupled receptor ligand
CC binding assays. The polypeptide fragments retain biological activity
CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC (see AAR4759-R48758, AAR50569-R50807 and AAR89189-R89195 for examples
CC of polypeptide fragments). The polypeptide fragments can be used in
CC compositions for treating subjects suffering from a pathology related to

CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
XX
SQ Sequence 317 AA;

Query Match 0.8%; Score 8; DB 15; Length 317;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 VAGSLSLV 707
| | | | | | | |
Db 7 vagslsllv 14

RESULT 11
AAM02661
ID AAM02661 standard; peptide; 317 AA.
XX
AC AAM02661;
XX
DT 12-NOV-1996 (first entry)
XX
DE G-protein coupled human m2 muscarinic acetylcholine receptor.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; CAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomegalovirus; serotonergic.
XX
OS Homo sapiens.
XX
PN US5508384-A.
XX
PD 16-APR-1996.
XX
PF 10-SEP-1992; 92US-0943236.
XX
PR 09-SEP-1993; 93US-0118270.
XX
PR 10-SEP-1992; 92US-0943236.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Murphy RB, Schuster DI;
XX
DR WPI: 1996-208785/21.
XX
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT for treating schizophrenia
XX
PS Disclosure: Column 49-52; 184pp; English.
XX
SQ Proteins AAM02657-W02720 represent a range of G-protein coupled receptor
CC (GPR) proteins selected from CAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomegaloviral and other GPR proteins. The receptor proteins
CC were used to design polypeptides, pref. based on the transmembrane
CC domains, for use in G-protein coupled receptor ligand binding assays.
CC The polypeptide fragments retain biological activity such as binding a
CC GPR ligand or modulating GPR ligand binding to a GPR (see
CC AAM02747-W02999 for examples of polypeptide fragments). The polypeptide
CC fragments can be used in compositions for treating subjects suffering
CC from a pathology related to a GPR abnormality e.g. a psychotic disorder
CC such as schizophrenia.
XX
SQ Sequence 317 AA;

Query Match 0.8%; Score 8; DB 17; Length 317;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 VAGSLSLV 707
| | | | | | | |


```

OS   Lycopersicon esculentum.
FH   Key                               Location/Qualifiers
FT   Peptide                          1..23
FT                                     /label= sig_peptide
FT   Protein                          24..1112
FT                                     /label= mature_peptide
XX
XX   WO9531564-A2.
XX
XX   23-NOV-1995.
XX
XX   11-MAY-1995;    95WO-GB01075.
XX
XX   07-APR-1995;    95GB-0007232.
XX   11-MAY-1994;    94GB-0009394.
XX   23-DEC-1994;    94WO-GB02812.
XX   31-MAR-1995;    95GB-0006658.
XX
XX   (GATS-) GATSBY CHARITABLE FOUND.
XX
XX   Hammond-Kosack KE, Jones DA, Jones JDG;
XX   WPI: 1996-010949/01.
XX   N-PSDB; AAT06306.
XX
XX   Increasing plant pathogen resistance by induction of variegation
XX   may lead to acquired resistance to a broad range of pathogens.
XX
XX   Disclosure; Page 84; 131pp; English.
XX
XX   AAR85298 is the tomato pathogen resistance gene Cf-2.1 protein. In a
XX   new method the Cf-2.1 gene is expressed highly in genetic constructs
XX   which may be used to impart a broad range of pathogen resistance, by
XX   induction of variegation, to transgenic plants (or parts or propagules
XX   of plants) containing such constructs. Cf-2.1 imparts resistance to
XX   the disease caused by the leaf mould fungal pathogen Cladosporium
XX   fulvum. C.fulvum contains avirulence (Avr) genes that confer
XX   recognition by plants containing Cf-genes, leading to the activation
XX   of host defence mechanisms to attack the disease.
XX
XX   Sequence 1112 AA;
XX
XX   Query Match                               0.8%; Score 8; DB 17; Length 1112;
XX   Best Local Similarity 100.0%; Pred. No. 99;
XX   Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY   758 SIGNLKNL 765
DB   330 signlnkl 337
XX
XX   RESULT 15
XX   AAR85299
XX   ID AAR85299 standard; Protein; 1112 AA.
XX
XX   AAR85299;
XX
XX   14-APR-1996 (first entry)
XX
XX   Tomato pathogen resistance gene Cf-2.2 protein.
XX
XX   Pathogen resistant; Cf-2.2; tomato; C.fulvum; Avr 4; Avr 9; fungal;
XX   leaf mould; variegation.
XX
XX   Lycopersicon esculentum.
XX
XX   Key                               Location/Qualifiers
XX   FH   Key                          1..23
XX   FT   Region                       /label= signal_peptide
XX   FT   Region                       24..1112

```

```

FT   /label= mature_peptide
XX
XX   WO9531564-A2.
XX
XX   23-NOV-1995.
XX
XX   11-MAY-1995;    95WO-GB01075.
XX
XX   07-APR-1995;    95GB-0007232.
XX   11-MAY-1994;    94GB-0009394.
XX   23-DEC-1994;    94WO-GB02812.
XX   31-MAR-1995;    95GB-0006658.
XX
XX   (GATS-) GATSBY CHARITABLE FOUND.
XX
XX   Hammond-Kosack KE, Jones DA, Jones JDG;
XX   WPI: 1996-010949/01.
XX
XX   Increasing plant pathogen resistance by induction of variegation
XX   may lead to acquired resistance to a broad range of pathogens.
XX
XX   Disclosure; Page 84-85; 131pp; English.
XX
XX   AAR85299 is the tomato pathogen resistance gene Cf-2.2 protein. In a
XX   new method the Cf-2.2 gene is expressed highly in genetic constructs
XX   which may be used to impart a broad range of pathogen resistance, by
XX   induction of variegation, to transgenic plants (or parts or propagules
XX   of plants) containing such constructs. Cf-2.2 imparts resistance to
XX   the disease caused by the leaf mould fungal pathogen Cladosporium
XX   fulvum. C.fulvum contains avirulence (Avr) genes that confer
XX   recognition by plants containing Cf-genes, leading to the activation
XX   of host defence mechanisms to attack the disease.
XX
XX   Sequence 1112 AA;
XX
XX   Query Match                               0.8%; Score 8; DB 17; Length 1112;
XX   Best Local Similarity 100.0%; Pred. No. 99;
XX   Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY   758 SIGNLKNL 765
DB   330 signlnkl 337
XX
XX   RESULT 16
XX   AAB42264
XX   ID AAB42264 standard; Protein; 2434 AA.
XX
XX   AAB42264;
XX
XX   08-FEB-2001 (first entry)
XX
XX   Human ORFX ORF2028 polypeptide sequence SEQ ID NO:4056.
XX
XX   Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
XX   vulnerary; antiparasitic; antiparkinsonian; noctropic; neuroprotective;
XX   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiact;
XX   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX   hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX   antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX   neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX   cholesterol ester storage; systemic lupus erythematosus; infection;
XX   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX   thrombosis; contraceptive.
XX
XX   Homo sapiens.
XX

```

```

PN WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, leach M;
XX
XX WPI: 2000-602362/57.
XX N-PSDB: AAC76473.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 3239-3244; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX osteoprotective; anticonvulsant; antidiabetic; immunosuppressive;
XX immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;
XX antihypertensive; antineoplastic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, anti-inflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX SQ Sequence 2434 AA;

```

```

XX
XX OS Homo sapiens.
XX
XX PN WO200153453-A2.
XX
XX PD 26-JUL-2001.
XX
XX PF 23-DEC-2000; 2000WO-US34960.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 30-NOV-2000; 2000US-0250583.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
XX Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Drmanac RT;
XX
XX WPI: 2001-488707/53.
XX N-PSDB: AAB90103.
XX
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
XX for treating e.g. cancer and immune deficiency disorders -
XX
XX Claim 10; Page 622-627; 648pp; English.
XX
XX The present sequence is one of 251 novel human polypeptides encoded
XX by a bone marrow-expressed polynucleotide. The polynucleotide and the
XX polypeptide encoded by it are useful in the treatment of various
XX immune deficiencies and disorders. The deficiencies and disorders may
XX be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
XX infection, or may result from an autoimmune disorder, a coagulation
XX disorder (e.g. hemophilia), inhibition of tumour cell proliferation,
XX suppression of an inflammatory response or treatment of a nervous
XX system disorder such as Alzheimer's disease. Detection of the presence
XX or increased expression of the polynucleotide or the protein it
XX encodes is useful for the diagnosis and/or prognosis of one
XX or more types of cancer. The polynucleotide and polypeptide can be
XX used as nutritional sources or supplements and in the screening of
XX chemical compounds as potential drugs.
XX
XX SQ Sequence 2594 AA;

```

```

Query Match 0.8%; Score 8; DB 21; Length 2434;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 895 SLSSLLKH 902
DB 14 SLSSLLKH 21

```

```

RESULT 17
AAB00984
ID AAB00984 standard; Protein; 2594 AA.
XX
XX AAB00984;
XX
XX 01-OCT-2001 (first entry)
XX
XX Human bone marrow protein, SEQ ID NO: 485.
XX
XX Human bone marrow; anti-inflammatory; cytoskeletal; neuroprotective;
XX antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
XX immunosuppressive; gene therapy; cytokine cell proliferation;
XX cell differentiation modulator; immune disorder; infection; cancer;
XX human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.

```

```

Query Match 0.8%; Score 8; DB 22; Length 2594;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 895 SLSSLLKH 902
DB 190 SLSSLLKH 197

```

```

RESULT 18
AAB66079
ID AAB66079 standard; Peptide; 22 AA.
XX
XX AAB66079;
XX
XX 30-MAR-2001 (first entry)
XX
XX Human INTERCEPT 296 transmembrane domain #5.
XX
XX TANGO protein; INTERCEPT protein; neurological disorder;
XX central nervous system; focal brain disorder; bipolar affective disorder;
XX global diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
XX senile dementia; Huntington's disease; amyotrophic lateral sclerosis;

```

KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KM neuropsychiatric; psychoactive substance use; anxiety.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PE 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
XX
PS Claim 8; Page 333; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
CC
XX
SQ Sequence 22 AA:

Query Match 0.7%; Score 7; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 LTTGLIC 428
Db 6 lttgllc 12
|||||
6 lttgllc 12

RESULT 19
AAB38032
ID AAB38032 standard; Peptide; 42 AA.
XX
AC AAB38032;
XX
DT 31-JAN-2001 (first entry)
XX
DE Fragment of human secreted protein encoded by gene 8 clone HBXFL29.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200055371-A1.
XX
PD 21-SEP-2000.
XX

PF 16-MAR-2000; 2000WO-US06783.
XX
PR 18-MAR-1999; 99US-0125055.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Edner R, Rosen CA, Shi Y, Birse C, Florence K;
XX Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX WPI; 2000-594448/56.
XX
DR
XX
PT New nucleic acid molecules encoding 27 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
XX
PS Disclosure; Page 22; 453pp; English.
XX
CC Sequences AAB37984-B38019 represent the amino acid sequences of 27
CC human secreted proteins encoded by the genes AAC69084-C69119. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC
XX
SQ Sequence 42 AA:

Query Match 0.7%; Score 7; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 VAGSLSL 706
Db 14 vagslsl 20
|||||
14 vagslsl 20

RESULT 20
AAY43097
ID AAY43097 standard; peptide; 46 AA.
XX
AC AAY43097;
XX
DT 17-JAN-2000 (first entry)
XX
DE Human integrin beta 2 subunit amino acid sequence.
XX
KW Integrin; beta 2 subunit; GP IIb; platelet aggregation; tyrosine
KW phosphorylation; thrombotic response; angiogenesis; tumour metastasis;
KW inflammation; transgenic animal.
XX
OS Homo sapiens.
XX
PN WO9953032-A1.
XX
PD 21-OCT-1999.
XX
PE 15-APR-1999; 99WO-US08285.
XX
PR 15-APR-1998; 98US-0115516.
XX
PA (CORT-) COR THERAPEUTICS INC.
XX
PI Law DA, Phillips DR;

XX
DR WPI; 1999-620415/53.
XX
PT Transgenic non-human mammals expressing mutant GP Iiia (integrin beta
subunit) gene for, e.g. studying platelet function -
XX
PS Disclosure: Fig 2; 37pp; English.

CC This is the human integrin beta 2 subunit amino acid sequence. Platelet
CC aggregation induced by a number of agonists, results in the
CC phosphorylation of tyrosine residues on the integrin beta 3 cytoplasmic
CC tail. The cytoplasmic domains of beta 1, beta 2 and beta 3 are important
CC for normal integrin/cytoskeleton interactions. The invention relates to a
CC non-human mammal containing a mutant GP Iiia gene where at least one of
CC the two cytoplasmic tyrosine residues of the gene has been replaced with
CC a non-tyrosine residue. Comparing a characteristic between two mammals of
CC the same species, where one mammal has a wild-type GP Iiia gene and the
CC other mammal has a mutant GP Iiia gene as above, can lead to the
CC determination of the effect of an agent on a characteristic of a mammal
CC that is attributable to the expression of the GP Iiia. The transgenic
CC mammals have the endogenous GP Iiia (an integrin beta subunit) gene
CC replaced with an altered or mutant GP Iiia gene. The platelets in the
CC blood of the resultant transgenic mammals cannot undergo tyrosine
CC phosphorylation to the extent that it occurs in wild-type mammals. They
CC therefore provide a model for assessing the importance of the
CC phosphorylation reaction for platelet function. The effect of mutant
CC GP Iiia on biological processes other than platelet formation can also be
CC studied. The transgenic non-human mammal may be used to compare the
CC bleeding time, thrombotic responses, angiogenesis, tumour metastasis or
CC inflammation between two mammals. The transgenic animals can also be used
CC to identify agents that modulate platelet aggregation or other effects
CC that are mediated by integrin signalling pathways.

XX
SQ Sequence 46 AA;

Query Match 0.7%; Score 7; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 LTHSDI 804
| | | | | | | |
Db 3 lthsd1 9

RESULT 21

AAM32656 standard; Protein: 59 AA.

XX
AC AAM32656;

XX
DT 17-OCT-2001 (first entry)

XX
DE Peptide #6693 encoded by probe for measuring placental gene expression.

XX
KM Probe: microarray; human; placenta; antenatal diagnosis;

XX
KM genetic disorder.

OS Homo sapiens.

XX
PN WO200157272-A2.

XX
PD 09-AUG-2001.

XX
PF 30-JAN-2001; 2001WO-US00663.

XX
PR 04-FEB-2000; 2000US-0180312.

XX
PR 26-MAY-2000; 2000US-0207456.

XX
PR 30-JUN-2000; 2000US-0608408.

XX
PR 03-AUG-2000; 2000US-0632366.

XX
PR 21-SEP-2000; 2000US-0234687.

XX
PR 27-SEP-2000; 2000US-0236359.

XX
PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-48897/53.

XX
PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 32925; 654pp; English.

XX
CC The present invention relates to single exon nucleic acid probes (SENP;
see AAI31315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
human genetic disorders.

XX
SQ Sequence 59 AA;

Query Match 0.7%; Score 7; DB 22; Length 59;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 AQDLKDL 108
| | | | | | | |
Db 25 aqdlkdl 31

RESULT 22

AAM20810 standard; Protein: 70 AA.

XX
AC AAM20810;

XX
DT 12-OCT-2001 (first entry)

XX
DE Peptide #7244 encoded by probe for measuring cervical gene expression.

XX
KM Probe: human; microarray; gene expression; cervical epithelial cell;

XX
KM cervical cancer.

OS Homo sapiens.

XX
PN WO200157278-A2.

XX
PD 09-AUG-2001.

XX
PF 30-JAN-2001; 2001WO-US00670.

XX
PR 04-FEB-2000; 2000US-0180312.

XX
PR 26-MAY-2000; 2000US-0207456.

XX
PR 30-JUN-2000; 2000US-0608408.

XX
PR 03-AUG-2000; 2000US-0632366.

XX
PR 21-SEP-2000; 2000US-0234687.

XX
PR 27-SEP-2000; 2000US-0236359.

XX
PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI; 2001-48897/53.

XX
PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 25636; 487pp; English.

XX
CC The present invention relates to human single exon nucleic acid probes
(SENP; see AAI10068-AA128459). The present sequence is a peptide encoded

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 585 YINSNI 591
 |||||
 Db 39 yinsgnl 45

RESULT 25
 AAY34819
 ID AAY34819 standard; Protein; 100 AA.

XX
 AC AAY34819;

DT 13-SEP-1999 (first entry)

DE Amino acid sequence of a Chlamydia pneumoniae protein.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

XX WO927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

PS Page 785; Disclosure; 1912pp; English.

CC AAY3584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see AAY3584-Y35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

XX Sequence 100 AA;

Query Match 0.7%; Score 7; DB 20; Length 100;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGLK 788
 |||||
 Db 15 klaeglk 21

RESULT 26

AAB57134
 ID AAB57134 standard; Protein; 100 AA.

XX AAB57134;

DT 13-MAR-2001 (first entry)

XX Human prostate cancer antigen protein sequence SEQ ID NO:1712.

DE Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.

XX Homo sapiens.

XX WO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-0505988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587513/55.

XX N-PSDB: AAF16337.

PT Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -

XX Claim 11; Page 2185; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.

CC The prostate cancer antigens can have neuroprotective, cytostatic,

CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,

CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome

CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat

CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders, wounds, and infectious diseases. AAF15566 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present

XX invention.

XX Sequence 100 AA;

Query Match 0.7%; Score 7; DB 21; Length 100;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 MAMLIKL 239
 |||||
 Db 70 mamlikl 76

RESULT 27

AAG01257
 ID AAG01257 standard; Protein; 105 AA.

XX AAG01257;

XX 06-OCN-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 5338.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX OS Homo sapiens.
 XX OS EP1033401-A2.
 XX PD 06-SEP-2000.
 XX PF 21-FEB-2000; 2000EP-0200610.
 XX PR 26-FEB-1999; 99US-0122487.
 XX PA (GEST) GENSET.
 XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX DR WPI: 2000-500381/45.
 XX DR N-PSDB; AAC01263.
 XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX XX
 XX PS Claim 13; SEQ ID 5338; 71pp + CD-ROM; English.
 XX CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 CC XX
 SQ Sequence 105 AA:
 Query Match 0.7%; Score 7; DB 21; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 782 KLAEGLK 788
 DB 91 klaeglk 97
 RESULT 28
 AAM42424
 ID AAM42424 standard; Protein; 118 AA.
 XX AC AAM42424;
 XX DT 22-JUN-1998 (first entry)
 XX DE Branched chain 2-oxo-acid dehydrogenase complex E2 subunit.
 XX KW Branched chain 2-oxo-acid dehydrogenase complex; BCOADC-E2; antigen;
 KW autoantigen; primary biliary cirrhosis; autoimmune disease;
 KW liver disease; anti-mitochondrial antibody; ELISA; assay;
 KW diagnosis; immunotherapy; vaccine.
 XX OS Bos taurus.
 XX PN W09749720-A1.
 XX PD 31-DEC-1997.
 XX PF 24-JUN-1997; 97WO-US11016.
 XX DR

PR 24-JUN-1996; 96US-0014719.
 XX PA (COPP/) COPPEL R.
 XX PA (GERS/) GERSHWIN E.
 XX PA (LEUNG/) LEUNG P.
 XX PI Coppel R, Gershwin E, Leung P;
 XX DR WPI: 1998-077105/07.
 XX DR N-PSDB; AAV03371.
 XX PT New fusion protein for detecting autoimmune liver disease -
 PT comprising the E2 subunit of pyruvate dehydrogenase, branched chain
 PT 2-oxo-acid dehydrogenase and 2-oxo glutarate dehydrogenase complexes
 XX XX
 XX PS Claim 6; Fig 4; 53pp; English.
 XX CC This protein comprises the lipolic acid binding domain E2 of bovine
 CC branched chain 2-oxo-acid dehydrogenase complex (BCOADC-E2). A
 CC claimed recombinant fusion protein (see AAM42427) comprises the E2
 CC subunit of the pyruvate dehydrogenase complex (PDC-E2, see AAM42425),
 CC the E2 subunit of BCOADC, and the E2 subunit of 2-oxo glutarate
 CC dehydrogenase complex (OGDC-E2, see AAM42426). The fusion protein
 CC can be used for detecting an anti-mitochondrial antibody (AMA) in a
 CC sample from a subject as indicative of an autoimmune liver disease,
 CC particularly primary biliary cirrhosis (PBC). The fusion protein
 CC can also be used in immunotherapy, e.g. production of vaccine. The
 CC fusion protein can detect antibodies to all 3 lipoyl domains of the
 CC E2 components with high sensitivity. BCOADC-E2 is recognised by
 CC approximately 60% of sera from PBC patients. Nucleic acid molecules
 CC (see AAV03370-73) encoding the fusion protein and its E2 components,
 CC as well as host-vector systems, are also claimed.
 CC XX
 SQ Sequence 118 AA:
 Query Match 0.7%; Score 7; DB 19; Length 118;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 801 LSDIGEG 807
 DB 8 lsdigeg 14
 RESULT 29
 AAY48392
 ID AAY48392 standard; Protein; 119 AA.
 XX AC AAY48392;
 XX DT 08-DEC-1999 (first entry)
 XX DE Human prostate cancer-associated protein 89.
 XX KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
 KW cancer; tissue specificity; human.
 XX OS Homo sapiens.
 XX PN DE19811194-A1.
 XX PD 16-SEP-1999.
 XX PF 10-MAR-1998; 98DE-1011194.
 XX PR 10-MAR-1998; 98DE-1011194.
 XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 XX DR WPI: 1999-519629/44.

Query Match 0.7%; Score 7; DB 20; Length 134;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 421 LITGL 427
Db 49 LITGL 55

RESULT 32
AAG54978
ID AAG54978 standard; Protein: 149 AA.

XX AAG54978;
AC
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 70357.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

```

PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 24-SEP-1999; 990S-0155486.
PR 28-SEP-1999; 990S-0155659.
PR 29-SEP-1999; 990S-0156458.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161922.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

```

Query Match 0.7%; Score 7; DB 21; Length 149;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 LGNUKLN 765
 Db 34 lgnlknl 40

```

RESULT 33
AAR58893
ID AAR58893 standard; Protein; 156 AA.
XX
AC AAR58893;
XX
DT 17-APR-1995 (first entry)
XX
DE Human-212 cadherin-related molecule.
XX
KM Cadherin; cell adhesion molecule.
XX
OS Homo sapiens.
XX
PN WO9414960-A.
XX
PD 07-JUL-1994.
XX
PF 23-DEC-1993; 93MO-US12588.
XX
PR 29-DEC-1992; 92US-0998003.
XX
PA (DOHE-) DOHENY EYE INST.
XX
PI Suzuki S.
XX
DR WPI; 1994-293849/36.
XX
DR N-PSDB; AAO68983.
XX
PT Polynucleotide sequences encoding new proto-cadherins - useful
PT for modulating natural binding and regulating activities.
XX
PS Example; Page 56; 114pp; English.
XX
CC Two regions of conserved AA sequence, one from the middle of the
CC third cadherin extracellular subdomain (EC-3) and the other from the
CC C-terminus of the fourth extracellular subdomain (EC-4) were
CC identified. The corresp. degenerate oligos (AAQ68949, AAO68950) were
CC designed for use as PCR primers. PCR was carried out on a rat brain
CC cDNA prep. Two major bands of about 450 bps and 130 bps were found.
CC The 450 bp band corresponded to the expected length between the two
CC primer sites, but the 130 bp band could not be predicted from any
CC of the previously identified cadherin sequences. The 450 bp and 130
CC bp bands were extracted and sequenced. Nineteen novel partial cDNA
CC clones were isolated. The DNA and deduced AA sequences of the
CC clones (including sequences corresp. to the PCR primers) are given
CC in AAO68951-Q68969 and AAR58860-R58878. Various cDNA fragments
CC structurally similar to the rat cDNAs were isolated from human,
CC mouse and Xenopus brain cDNA preps. and from Drosophila and
CC C. elegans whole body cDNA preps. by PCR using the above primers.
CC The DNA and deduced AA sequences of the resulting PCR fragments
CC (including sequences corresp. to the PCR primers) are given in
CC AAO68971, AAO68972-Q68994 and AAR58882-R58905 and AAR49143. Comparison
CC of the deduced AA sequences indicates a similarity, in particular,
CC there are three sets of clones that appear to be cross-species
CC homologues: RAT-218, MOUSE-322 and HUMAN-43; RAT-314, MOUSE-321 and
CC HUMAN-11; and MOUSE-326 and HUMAN-42.
XX
SQ Sequence 156 AA:

```

Query Match 0.7%; Score 7; DB 15; Length 156;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 LTINGLL 158
 Db 103 ltingll 109

RESULT 34

AAR87132
 ID AAR87132 standard; Peptide; 156 AA.
 AC AAR87132;
 XX
 DT 29-AUG-1996 (first entry)
 DE Protocadherin clone HUMAN-212.
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion; mouse;
 KM catenin; therapy; clone.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 118
 FT /note= "encoded by CAP"
 XX
 PM WO9600289-A1.
 XX
 PD 04-JAN-1996.
 XX
 PF 26-JUN-1995; 95MO-US08071.
 XX
 PR 27-JUN-1994; 94US-0268161.
 XX
 PA (DOHE-) DOHENY EYE INST.
 XX
 PI Suzuki S;
 XX
 DR WPI: 1996-068873/07.
 DR N-PSDB: AAT03607.
 XX
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PS pc5 - involved in cell-cell adhesion and regulation activities
 XX
 PS Example 2; Page 60; 146pp; English.
 XX
 CC AAR87135-R87137 represent partial fragments of the human protocadherin
 CC sequence. The cDNAs encoding these sequences were isolated after
 CC screening a human brain cDNA preparation with the primers shown in
 CC AAT03575 and AAT03576. The primers were constructed from portions of the
 CC amino acid sequences of the third and fourth extracellular domains of
 CC published cadherin sequences. The full length cDNA sequences encoding
 CC human protocadherins pc3 and pc4 are represented by AAT03572 and
 CC AAT03573, respectively. The cytoplasmic domain of cadherin interacts
 CC with the cytoskeleton through catenins and other cytoskeleton associated
 CC proteins. The cytoplasmic domain is not present in all cadherins, but in
 CC those which possess it, it is essential for the cadherins adhesive
 CC function. The cadherins which do not possess a cytoplasmic domain appear
 CC to function via a different method from those with a cytoplasmic domain.
 CC These protein sequences are involved in cell-cell adhesion. These
 CC sequences may have regulatory functions in the cell, as well as the
 CC cell-cell adhesive properties. Antibodies produced against these
 CC sequences are useful for modulating the binding activity of these
 CC protocadherins, and can be used therapeutically.
 CC
 XX Sequence 156 AA:
 SQ
 Query Match 0.7%; Score 7; DB 17; Length 156;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 152 LFTNGLL 158
 DB 103 LFTNGLL 109
 RESULT 35
 AAB25353
 ID AAB25353 standard; Protein; 162 AA.
 XX

AC AAB25353;
 XX
 DT 27-NOV-2000 (first entry)
 DE Pinus radiata cell signalling involved protein SEQ ID NO:672.
 XX
 KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
 KM plant cell signalling; modulation; transgenic plant; pathogen; growth;
 KW environmental change; development; cell proliferation; differentiation;
 KM elongation; survival; disease resistance; nutrient metabolism.
 XX
 OS Pinus radiata.
 XX
 PN WO200042171-A1.
 XX
 PD 20-JUL-2000.
 XX
 PF 11-JAN-2000; 2000MO-US00724.
 XX
 PR 12-JAN-1999; 99US-0228986.
 XX
 PR 01-NOV-1999; 99US-0162866.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Strabala TJ, Nieuwenhuizen NJ;
 XX
 DR WPI: 2000-476052/41.
 XX
 PT Isolated polynucleotide encoding a polypeptide involved in cell
 PS signaling used for generating transgenic plants with modified responses
 XX to external signals -
 XX
 PS Claim 3; Page 307-308; 527pp; English.
 XX
 CC AAA97263 to AAA9736 and AAB25100 to AAB25570 represent polynucleotide
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences
 CC are involved in cell signalling. The polynucleotide and protein
 CC sequences can be used to modify the response of plant cells to external
 CC signals e.g. environmental changes or pathogens during the growth and
 CC development of a plant. They can be used to modify cell proliferation,
 CC differentiation, elongation and survival, resistance to disease and
 CC nutrient metabolism. Examples of modifications which can be produced are
 CC altered fruit ripening and senescence of leaves and flowers e.g. to
 CC delay senescence and prolong the life of cut flowers or enhance
 CC senescence of reproductive organs to engineer sterile plants. Other
 CC modifications can be used to delay senescence in selected cell types or
 CC organs providing fruit and vegetables which have a longer shelf life
 CC between harvest and consumption, or to decrease branching frequency in
 CC forest tree species giving long stretches of valuable knot-free clear
 CC wood which can be used in solid timber furniture and veneers.
 CC
 XX Sequence 162 AA:
 SQ
 Query Match 0.7%; Score 7; DB 21; Length 162;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 894 GSLSLL 900
 DB 59 GSLSLL 65
 RESULT 36
 AAU00288
 ID AAU00288 standard; Protein; 166 AA.
 AC AAU00288;
 XX
 DT 12-SEP-2001 (first entry)
 DE Human steroid hormone receptor encoded by gene HMO64.
 XX

XX Human: soluble human steroid hormone receptor; diagnostic; anaemia;
KM immune disease; human immunodeficiency virus; HIV; autoimmune disease;
KM rheumatoid arthritis; thyroiditis; diabetes mellitus; cancer; wound;
KM blood coagulation disease; inflammatory condition; asthma; infection;
KM gastrointestinal disorder; inflammatory bowel disease; ocular disorder;
KM central nervous system disorder; multiple sclerosis; Alzheimer's disease;
KM neurodegenerative disorder; Parkinson's disease; cardiovascular disorder;
KM chromosome identification; forensic science; molecular weight marker.
XX
OS Homo sapiens.
XX
PN WO200118049-A1.
XX
PD 15-MAR-2001.
XX
PF 07-SEP-2000; 2000MO-US24517.
XX
PR 09-SEP-1999; 99US-0152932.
PR 14-MAR-2000; 2000US-0189032.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI NI J, Shi Y, Ruben SM;
XX
DR WPI: 2001-183281/18.
DR N-PSDB; AAS01141.
XX
PT New nucleic acid encoding a soluble human steroid hormone receptor and
PT its associated polypeptides and antibodies, useful for the diagnosis,
PT prevention, treatment and amelioration of e.g. (auto)immune diseases
PT and cancers
XX
XX
PS Claim 11: Page 10; 241pp; English.
XX
XX The sequence represents the amino acid sequence of human steroid hormone
CC receptor encoded by gene HMHQR64. SHR polypeptides and polynucleotides
CC can be used to diagnose, prevent, treat or ameliorate diseases and
CC conditions including immune diseases (e.g. human immunodeficiency virus
CC (HIV) and anaemia), autoimmune diseases (e.g. rheumatoid arthritis,
CC thyroiditis and diabetes mellitus), blood coagulation diseases, wounds,
CC inflammatory conditions (e.g. asthma), gastrointestinal disorders (e.g.
CC inflammatory bowel disease), cancers, central nervous system disorders
CC (e.g. multiple sclerosis), neurodegenerative disorders (e.g. Parkinson's
CC disease and Alzheimer's disease), cardiovascular disorders, ocular
CC disorders and diseases caused by infectious agents (e.g. viruses such as
CC HIV, bacteria, fungi and parasites). Binding partners (antagonists and
CC agonists) antisense sequences and ribozymes are also useful for
CC preventing, treating and ameliorating the above diseases and disorders.
CC The polypeptides and polynucleotides can also be used to detect, diagnose
CC or monitor diseases and conditions associated with aberrant expression or
CC activity of SHR. The polynucleotides are also useful for chromosome
CC identification, as primers in polymerase chain reactions, identifying
CC individuals in forensic science and as molecular weight markers.
CC Antibodies to SHR are also useful as immunological probes for
CC differential identification of tissues and as molecular weight markers.
XX
SO Sequence 166 AA:

Query Match 0.7%; Score 7; DB 22; Length 166;
Best local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 699 GYAGSUS 705
DB 50 GYAGSIS 56

RESULT 37
AAG09451
ID AAG09451 standard; Protein: 175 AA.
XX

AC AAG09451;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 7992.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

```
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139811.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144003.
PR 16-JUL-1999; 99US-0144083.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144881.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145118.
PR 23-JUL-1999; 99US-0145214.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145291.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147202.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160811.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.7%; Score 7; DB 21; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 LSLVLT 710
Db 169 lslvlt 175

RESULT 38
AAG38090
ID AAG38090 standard; Protein; 175 AA.
XX
AC AAG38090;
XX
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46938.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
```


KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142820.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153756.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161358.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 0.7%; Score 7; DB 21; Length 175;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 704 LSLVST 710
 |||||
 DB 169 lslvst 175

RESULT 39

AAW44388
 ID AAW44388 standard; Protein: 180 AA.

AC AAW44388;
 XX

DT 10-JUN-1998 (first entry)
 XX

DE Vancomycin resistant Enterococcus faecium expression product.
 XX

KW Vancomycin resistant Enterococcus faecium; VRE; epitope; immunogen;
 KW methicillin resistant Staphylococcus aureus; MNSA; transporter protein.

OS Enterococcus faecium.
 XX

PN WO9801154-A2.
 XX

PD 15-JAN-1998.

```

XX 07-JUL-1997; 97WO-GB01830.
XX
XX 06-JUL-1996; 96GB-0014274.
XX
XX (NEUT-) NEUTEC PHARM PLC.
XX
XX Burnie JP, Matthews RC;
XX
XX WPI: 1998-100821/09.
XX N-PSDB; AAV15410.
XX
XX Treating and diagnosing bacterial and fungal infection with ABC
XX transporter protein - or neutralising or binding agents, and new
XX staphylococcal proteins, particularly for infections caused by drug
XX resistant Staphylococci and Enterococci
XX
XX Example: Page 36; 62pp; English.
XX
XX The present sequence represents a vancomycin resistant Enterococcus
XX faecium (VRE) expression product. The present invention describes the
XX therapeutic or diagnostic use of a bacterial or fungal ABC transporter
XX protein, or its immunogenic fragments, in human or veterinary medicine.
XX Binding agents, specific for an ABC transporter protein, are used in
XX standard binding assays to detect the ABC transporter protein, while
XX ABC transporter proteins are used similarly to detect specific
XX antibodies in patient sera, specifically for diagnosis of Staphylococci
XX and Enterococci infections. Neutralising agents specific for ABC
XX transporter proteins, specifically antibodies (Ab), are used to treat
XX such infections (optionally when coupled to a drug) while ABC
XX transporter proteins are used as immunogens to protect against
XX infection. Ab raised against ABC transporter proteins are effective
XX against multiple drug resistant strains of bacteria (specifically
XX vancomycin and/or methicillin resistant strains) for which no drug
XX therapy is available.
XX
XX Sequence 180 AA;
XX

```

Query Match 0.7%; Score 7; DB 19; Length 180;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180
 |||||
 DB 49 gkstllq 55

RESULT 40

AAG09450
 ID AAG09450 standard; Protein: 208 AA.

AC AAG09450;
 XX

DT 17-OCT-2000 (first entry)
 XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7391.
 XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.
 XX

PN EP1033405-A2.
 XX

PD 06-SEP-2000.
 XX

PF 25-FEB-2000; 2000EP-0301439.
 XX

PR 25-FEB-1999; 99US-0121825.
 XX

PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135553.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136382.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.

PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151338.
PR 01-SEP-1999; 990S-0151930.
PR 01-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.7%; Score 7; DB 21; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 704 LSLVST 710
|||||||
Db 202 lslvst 208

RESULT 41
AAG38089
ID AAG38089 standard; Protein; 208 AA.
XX
AC AAG38089;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 46937.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

```

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```

```

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          0.7%: Score 7; DB 21; length 208;
Best Local Similarity 100.0%: Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 LSLVLSST 710
Db 202 lslvlst 208

RESULT 42
AAU03694
ID AAU03694 standard; Protein; 211 AA.
XX
AC AAU03694;
XX
DT 12-SEP-2001 (first entry)
XX
DE Group B Streptococcus antigenic protein, ID-171.
XX
KW Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;
KW meningitis; neonate; antigenic; vaccine; infection; genital tract;
XX
OS Streptococcus agalactiae.
XX
PN WO200132882-A2.
XX
PD 10-MAY-2001.
XX
PF 07-SEP-2000; 2000WO-GB03437.
XX
PR 07-SEP-1999; 99GB-0021125.
XX
(MICR-) MICROBIAL TECHNIQS LTD.
XX
Le Page RWF, Wells JM, Hanniffy SB;
XX
WPI; 2001-316444/33.
XX
N-PSDB; AAS07111.
XX

New polypeptides derived from Streptococcus agalactiae are useful to
PT provide detection of, and vaccination against, Group B Streptococcus
PT infections, particularly to prevent infection in neonatals
XX
PS Claim 1; Fig 1; 178pp; English.
XX
CC AAU03601-AAU03722 represent Group B Streptococcus (Streptococcus
agalactiae) amino acid sequences of the invention. S. agalactiae is an

```

CC encapsulated bacterium which is a major pathogen of humans causing sepsis
CC and meningitis in neonates as well as adults. The S. agalactiae antigenic
CC polypeptides are used to vaccinate against Group B Streptococcus
CC infections, particularly to prevent infection in new born children
CC arising from the maternal genital tract. An immunogenic composition is
CC useful in the preparation of a medicament for the treatment or
CC prophylaxis of Group B Streptococcus infection. The invention does not
CC have the disadvantages of varied response rate associated with prior art
CC capsid polysaccharide vaccination against Group B Streptococcus.

XX Sequence 211 AA:

Query Match 0.7%; Score 7; DB 22; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 519 LGLSIK 525
|||||||
DB 162 LGLSIK 168

RESULT 43

AAG72854
ID AAG72854 standard; Protein: 215 AA.

XX AAG72854;

DT 30-JUL-2001 (first entry)

XX Human olfactory receptor data exploratorium sequence, SEQ ID NO: 2536.

DE Human: olfactory receptor; OR: primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation;
KW human olfactory receptor data exploratorium; HORDE.

XX Homo sapiens.

OS WO200127158-A2.

PN 19-APR-2001.

PD 06-OCT-2000; 2000WO-US27582.

PF 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

PA (DIGI-) DIGISCENTS.
PA (YEDA) YEDA RES & DEV CO LTD.

XX Belenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI: 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -

XX Example 6; Page 1716-1717; 1857pp; English.

XX The present sequence is a polypeptide from the human olfactory receptor
CC data exploratorium (HORDE). It was used as a query sequence in a database
CC search of olfactory receptor (OR)-like sequences. The invention relates
CC to isolated polynucleotides encoding polypeptides involved in olfactory
CC sensation. The polynucleotides can be used in screening for olfactory
CC agonists and antagonists. The methods allow for the determination of
CC primary scents and the identification of the odour receptors used to
CC detect these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents. This
CC enables the construction of a scent representation (also called a scent
CC fingerprint or scent profile), which may be used to re-create and edit
CC scents. Libraries of olfactory receptors are useful for determining the

CC interaction pattern of a composition with the receptors, and can be
CC used for determining differences in the olfactory faculties of different
CC individuals.

XX Sequence 215 AA:

Query Match 0.7%; Score 7; DB 22; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 486 TYSSLR 492
|||||||
DB 34 TYSSLR 40

RESULT 44

AAB79373
ID AAB79373 standard; Protein: 215 AA.

XX AAB79373;

DT 30-APR-2001 (first entry)

XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:262.

DE Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteologenic amino acid;
KW nonproteologenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carboxylate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
KW diagnosis; Corynebacterium diptheriae; evolutionary study.

OS Corynebacterium glutamicum.

OS WO200100844-A2.

PN 04-JAN-2001.

PD 23-JUN-2000; 2000WO-IB00943.

PF 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.

PR 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032180.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032230.

PR 09-JUL-1999; 99US-0143208.

PR 14-JUL-1999; 99DE-1032924.

PR 14-JUL-1999; 99DE-1032973.

PR 27-AUG-1999; 99DE-1033005.

PR 31-AUG-1999; 99US-0151572.

PR 03-SEP-1999; 99DE-1042076.

PR 03-SEP-1999; 99DE-1042079.

PR 03-SEP-1999; 99DE-1042086.

PR 03-SEP-1999; 99DE-1042087.

PR 03-SEP-1999; 99DE-1042088.

PR 03-SEP-1999; 99DE-1042095.

PR 03-SEP-1999; 99DE-1042123.

PR 03-SEP-1999; 99DE-1042125.

PA (BAD1) BASF AG.
 XX
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 XX WPI; 2001-061975/07.
 DR N-PSDB; AAF71490.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 515-516; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteoligogenic
 CC or nonproteoligogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SQ Sequence 215 AA;

Query Match 0.7%; Score 7; DB 22; Length 215;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSL 706
 |||||
 Db 118 VAGSLSL 124

RESULT 45
 AAM47019
 ID AAM47019 standard; Protein; 218 AA.
 XX
 AC AAM47019;
 XX
 DT 03-AUG-1998 (first entry)
 XX
 DE Arabidopsis thaliana SERK LRR homologous EST clone encoded protein.
 XX
 KM receptor kinase; apomixis; apomictic; seeds; production; embryos;
 KM plant breeding; leucine-rich repeat.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO9743427-A1.
 XX
 PD 20-NOV-1997.
 XX
 PF 13-MAY-1997; 97WO-EP02443.
 XX
 PR 14-MAY-1996; 96GB-0010044.
 XX
 PA (NOVS) NOVARTIS AG.
 XX
 PI De Vries SC, Hecht VFG, Schmidt EDL, Van Holst GJ;
 XX WPI; 1998-086529/08.
 DR

DR N-PSDB; AAV06587.
 XX
 PT Production of apomictic seeds - useful in plant breeding
 XX
 PS Disclosure; Pages 77-78; 123pp; English.
 XX
 CC The sequence is that encoded by an EST clone showing high homology to
 CC SERK LRR (leucine-rich repeat) sequences.
 XX
 SQ Sequence 218 AA;

Query Match 0.7%; Score 7; DB 19; Length 218;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 759 LGNLEKNL 765
 |||||
 Db 114 LGNLEKNL 120

RESULT 46
 AAM47022
 ID AAM47022 standard; Protein; 218 AA.
 XX
 AC AAM47022;
 XX
 DT 03-AUG-1998 (first entry)
 XX
 DE Arabidopsis thaliana SERK LRR homologous EST clone encoded protein.
 XX
 KM receptor kinase; apomixis; apomictic; seeds; production; embryos;
 KM plant breeding; leucine-rich repeat.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO9743427-A1.
 XX
 PD 20-NOV-1997.
 XX
 PF 13-MAY-1997; 97WO-EP02443.
 XX
 PR 14-MAY-1996; 96GB-0010044.
 XX
 PA (NOVS) NOVARTIS AG.
 XX
 PI De Vries SC, Hecht VFG, Schmidt EDL, Van Holst GJ;
 XX WPI; 1998-086529/08.
 DR N-PSDB; AAM47022.
 XX
 PT Production of apomictic seeds - useful in plant breeding
 XX
 PS Disclosure; Pages 89-90; 123pp; English.
 XX
 CC The sequence is that encoded by an EST clone showing high homology to
 CC SERK LRR (leucine-rich repeat) sequences.
 XX
 SQ Sequence 218 AA;

Query Match 0.7%; Score 7; DB 19; Length 218;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 759 LGNLEKNL 765
 |||||
 Db 114 LGNLEKNL 120

RESULT 47
 AAM47018
 ID AAM47018 standard; Protein; 218 AA.
 DR

XX AAW47018;
AC 03-AUG-1998 (first entry)
XX
DT Arabidopsis thaliana SERK LRR homologous EST clone encoded protein.
XX
DE Arabidopsis thaliana SERK LRR homologous EST clone encoded protein.
XX
KW receptor kinase; apomixis; apomictic; seeds; production; embryos;
XX plant breeding; leucine-rich repeat.
XX
OS Arabidopsis thaliana.
XX
PN W09743427-A1.
XX
PD 20-NOV-1997.
XX
PF 13-MAY-1997; 97WO-EP02443.
XX
PR 14-MAY-1996; 96GB-0010044.
XX
PA (NOVS) NOVARTIS AG.
XX
PI De Vries SC, Hecht VFC, Schmidt EDL, Van Holst GJ;
XX WPI: 1998-086529/08.
DR N-PSDB: AAV06586.
XX
PT Production of apomictic seeds - useful in plant breeding
XX
PS Disclosure: Pages 73-74, 123pp; English.
XX
CC The sequence is that encoded by an EST clone showing high homology to
CC SERK LRR (leucine-rich repeat) sequences.
XX
SQ Sequence 218 AA;

Query Match 0.7%; Score 7; DB 19; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 LGNIKNL 765
DB 114 LGNIKNL 120

RESULT 48
AAG07961
ID AAG07961 standard; Protein; 218 AA.
XX
AC AAG07961;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5312.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.7%; Score 7; DB 21; Length 218;
Best local Similarity 100.0%; Pred. No. 2,2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 LGNKNL 765
Db 114 lgnlknl 120

RESULT 49
AAG90022
ID AAG90022 standard; Protein; 218 AA.
XX
AC AAG90022;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 3776.
XX
KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX Corynebacterium glutamicum.
OS
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI, 2001-376931/40.
DR N-PSDB; AAH65241.
XX
PT Novel polynucleotides derived from Corynebacterium, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
PS Claim 17; SEQ ID NO: 3776; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium, and identifying a homologue of a gene derived
CC from corynebacterium. Corynebacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 218 AA;

Query Match 0.7%; Score 7; DB 22; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 STLLQRI 182
|||
Db 124 stllqrl 130

RESULT 50
AAW38486
ID AAW38486 standard; Protein: 220 AA.

XX
AC AAW38486;

DT 06-NOV-1998 (first entry)

DE S. pneumoniae cell division protein FTSA.

XX
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis.
XX
OS Streptococcus pneumoniae.

XX
FN W09743303-A1.

PD 20-NOV-1997.

PF 14-MAY-1997; 97WO-US07950.

PR 14-MAY-1996; 96US-0017670.

PA (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

PI Stodola RK;

DR WPI: 1998-008793/01.

DR N-PSDB; AAT98554.

PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
XX
PS Claim 12; Pages 276-277; 483pp; English.

CC This sequence represents a Streptococcus pneumoniae protein that, based
CC on homology with a Bacillus subtilis protein, is a cell division
CC protein FTSA, and is encoded by a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
XX
SQ Sequence 220 AA;

Query Match 0.7%; Score 7; DB 19; Length 220;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGIK 788
|||
Db 136 klaegik 142

RESULT 51
AAW47020
ID AAW47020 standard; Protein: 220 AA.

XX
AC AAW47020;

DT 03-AUG-1998 (first entry)

DE Arabidopsis thaliana SERK LRR homologous EST clone encoded protein.

XX
KW receptor kinase; apomixis; apomictic; seeds; production; embryos;
KW plant breeding; leucine-rich repeat.
XX
OS Arabidopsis thaliana.

XX
FN W09743427-A1.

PD 20-NOV-1997.

PF 13-MAY-1997; 97WO-EP02443.

PR 14-MAY-1996; 96GB-0010044.

PA (NOVS) NOVARTIS AG.

PI De Vries SC, Hecht VFG, Schmidt EDL, Van Holst GJ;

DR WPI: 1998-086529/08.

DR N-PSDB; AAW47020.

PT Production of apomictic seeds - useful in plant breeding

XX
PS Disclosure; Pages 81-82; 123pp; English.

CC The sequence is that encoded by an EST clone showing high homology to
CC SERK LRR (leucine-rich repeat) sequences.
XX
SQ Sequence 220 AA;

Query Match 0.7%; Score 7; DB 19; Length 220;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 759 LGNKNL 765
| | | | | | | |
Db 116 lgnlkn1 122

RESULT 52

AAB51901
ID AAB51901 standard; Protein; 222 AA.

AC AAB51901;

DT 16-FEB-2001 (first entry)

DE Human secreted protein sequence encoded by gene 21 SEQ ID NO:134.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KM antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KM neotropic; neuroprotective; antibacterial; virucide; fungicide;
KM optalmalogical; vulnerray; autoimmune disease; Rheumatoid arthritis;
KM hyperproliferative disorders; cancer; cardiovascular disorder;
KM cardiac arrest; cerebrovascular disorder; nervous system disorder;
KM Alzheimer's disease; ocular disorder; wound healing; skin aging.

OS Homo sapiens.

XX WO200061626-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000MO-US09066.

PR 09-APR-1999; 99US-0128698.

PR 20-JAN-2000; 2000US-0176926.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE-) ROSEN C A.

PI Rosen CA, Ruben SM, Komatsoulis G;

DR WPI: 2000-619227/59.

DR N-PSDB: AAC93499.

PT New nucleic acid molecules encoding 49 human secreted proteins for

PT diagnosing, preventing or ameliorating medical conditions and used for

PT food additives or preservatives -

PS Disclosure; Page 58; 516pp; English.

XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding

CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -

CC AAB51927 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and

CC proteins have activities dependent on the tissues and cells in which they

CC are expressed. Examples of their activities include immunosuppressive;

CC antirheumatic; antirheumatic; antiproliferative; cytostatic; cardiant;

CC vasotropic; cerebroprotective; neotropic; neuroprotective; antibacterial;

CC virucide; fungicide; optalmalogical; and vulnerray. The secreted

CC proteins, polynucleotides, antagonists and agonists may be useful in

CC treating, preventing and/or diagnosing diseases and disorders such as

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,

CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,

CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
CC used in the isolation and characterisation of the proteins and
CC polynucleotides of the invention.

SQ Sequence 222 AA;

Query Match 0.7%; Score 7; DB 21; Length 222;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 ILGAFEG 932
| | | | | | | |

Db 159 llgafg 165

RESULT 53

AAB28796
ID AAB28796 standard; Protein; 231 AA.

AC AAB28796;

DT 13-FEB-2001 (first entry)

DE Human hydrolase-like molecule 7 protein.

XX Hydrolase-like molecule; human; cell proliferation disorder;
KM autoimmune; cancer; AIDS; acquired immune deficiency syndrome.

OS Homo sapiens.

XX US6132964-A.

PD 17-OCT-2000.

PF 06-FEB-1998; 98US-0013881.

PR 06-FEB-1998; 98US-0013881.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Hillman JL, Guegler KD, Shah P, Lal P, Corley NC;

DR WPI: 2001-006133/01.

DR N-PSDB: AAC60229.

PT New human hydrolase-like molecules (HILMs) and polynucleotides encoding

PT the HILMs, useful for diagnosing, treating or preventing cell

PT proliferation (e.g. bone cancer) or autoimmune disorders(e.g. AIDS or

PT asthma) -

PS Claim 1; Column 53-54; 38pp; English.

CC The present invention relates to isolated and purified cDNA encoding a

CC human hydrolase-like molecule (HILM), designated HILM-1 to HILM-8. The

CC HILM DNAs and polypeptides are useful for diagnosing, treating or

CC preventing cell proliferation disorders and autoimmune disorders. Cell

CC proliferation disorders include cancers, autoimmune disorders include

CC AIDS (acquired immune deficiency syndrome). The present sequence

CC is a HILM protein of the invention.

SQ Sequence 231 AA;

Query Match 0.7%; Score 7; DB 22; Length 231;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 FGALTAE 296
| | | | | | | |

Db 183 fgaltae 189

RESULT 54
 AAB43179 standard; Protein; 265 AA.
 ID AAB43179;
 AC AAB43179;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF2943 polypeptide sequence SEQ ID NO:5886.
 XX
 KW Human: open reading frame; ORFX; detection; cytosstatic; hepatotropic;
 KW vulnery; antiporiatic; antiparkinsonian; noctropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 PN WO200058473-AZ.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR N-PSDB: AAC77388.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 5059-5060; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiporiatic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypoid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hyperension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 265 AA;
 Query Match 0.7%; Score 7; DB 21; Length 265;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 926 ILGAFFG 932
 |||||
 DB 196 ILGAFFG 202
 RESULT 55
 AAY72607
 ID AAY72607 standard; Protein; 265 AA.
 XX
 AC AAY72607;
 XX
 DT 02-MAY-2001 (first entry)
 DE Human Electron Transfer Protein, ETRN-5.
 XX
 KW Human; electron transfer protein; ETRN-5; cytosstatic; immunosuppressive;
 KW dermatological; nephrotrophic; thyromimetic; neuroprotective; allergy;
 KW antituber; therapy; cell proliferative disorder; arteriosclerosis; gout;
 KW atherosclerosis; psoriasis; reproductive disorder; infertility; anaemia;
 KW immune disorder; cancer; adenocarcinoma; leukaemia; gene therapy;
 KW autoimmune thyroiditis; Crohn's disease; rheumatoid arthritis; asthma;
 KW diabetes mellitus; Acquired Immune Deficiency Syndrome; AIDS;
 KW Addison's disease.
 XX
 OS Homo sapiens.
 PN Key Location/Qualifiers
 FT Region 38..255
 FT /label= ubiquinone_biosynthesis_protein_motif
 XX
 PN WO200105969-A2.
 PD 25-JAN-2001.
 XX
 PF 12-JUL-2000; 2000WO-US19036.
 PR 14-JUL-1999; 99US-0143816.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lal P, Tang YT, Yue H, Baughn MR, Lu DAM;
 XX
 DR WPI: 2001-147342/15.
 DR N-PSDB: AAD02598.
 XX
 PT Human electron transfer proteins (ETRN) useful for diagnosing,
 PT treating, preventing disorders associated with abnormal expression of
 PT ETRN, -
 XX
 PS Claim 1; Page 81; 84pp; English.
 XX
 CC The present sequence is human electron transfer protein-5 (ETRN-5).
 CC The ETRN-5 cDNA with Incyte clone ID 515362 is obtained from
 CC OVADIR704 cDNA library. ETRN are used in the diagnosis,
 CC prevention and treatment of cell proliferative disorders (actinic
 CC keratosis, arteriosclerosis, atherosclerosis, burns, cirrhosis,
 CC hepatitis, mixed connective tissue disease (MCTD), psoriasis,
 CC myelofibrosis); cancers (adenocarcinoma, leukaemia, lymphoma, melanoma);
 CC reproductive disorders (prolactin disorder, infertility, tubal disease,
 CC disruption of oestrous cycle, disruption of menstrual cycle, prostatitis,
 CC ectopic pregnancy, spermatogenesis, cancer of testis, prostate) and
 CC immune response disorders (acquired immunodeficiency syndrome (AIDS),
 CC Addison's disease, adult respiratory distress syndrome, allergies,
 CC asthma, amyloidosis, anaemia, autoimmune hemolytic anaemia, autoimmune
 CC thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal

dystrophy (APECED), Crohn's disease, atopic dermatitis, gout, diabetes mellitus, rheumatoid arthritis or ulcerative colitis, Goodpasture's syndrome, Hashimoto's thyroiditis, osteoarthritis, multiple sclerosis, osteoporosis). ETRN or its immunogenic fragments are useful in screening libraries and in drug screening assays. ETRN polynucleotides may also be useful in somatic or germ-line gene therapy.

Sequence 265 AA;

Query Match 0.7%; Score 7; DB 22; Length 265;

Best Local Similarity 100.0%; Pred. No. 2.7e+02; Mismatches 0; Indels 0; Gaps 0;

OY 926 ILGAFPG 932

Db 196 IIGAFPG 202

RESULT 56

AA01298

ID AA01298 standard; Protein; 277 AA.

AC AA01298;

DT 04-OCT-2001 (first entry)

DE P501S-specific T cell clone 4E5 Va chain T cell receptor amino acid.

KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;

KW cytostatic; gene therapy; metastasis.

OS Homo sapiens.

PN WO200151633-A2.

PD 19-JUL-2001.

PF 16-JAN-2001; 2001WO-US01574.

PR 14-JAN-2000; 2000US-0483672.

PA (CORI-) CORIXA CORP.

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;

PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM;

PI Wang A, Meagher MJ;

DR WPI; 2001-425873/45.

PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -

PS Example 24; Page 525-526; 543pp; English.

CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used to treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AA093357 to AA093944 and AA01115 to
CC AA01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.

Sequence 277 AA;

Query Match 0.7%; Score 7; DB 22; Length 277;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;

OY 895 SLSSLLK 901

Db 2 SLSSLLK 8

RESULT 57

AA091261

ID AA091261 standard; Protein; 282 AA.

AC AA091261;

DT 26-SEP-2001 (first entry)

DE C glutamylum protein fragment SEQ ID NO: 5015.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

PA (KYOWA) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR N-PSDB; AA066480.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 5015; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

Sequence 282 AA;

Query Match 0.7%; Score 7; DB 22; Length 282;

Best Local Similarity 100.0%; Pred. No. 2.9e+02; Mismatches 0; Indels 0; Gaps 0;

OY 379 KGVAASD 385

|||||||

DB	23	kgyaasd	29	
RESULT	58			
AAW36004				
ID	AAW36004	standard; Protein; 283 AA.		
XX				
AC	AAW36004;			
XX				
DT	03-MAR-1998	(first entry)		
XX				
DE	Human Fchd545 gene product.			
XX				
KW	Fchd545 gene; differential expression; endothelial cell; human;			
KW	shear stress; cardiovascular disease; atherosclerosis; ischaemia;			
KW	reperfusion; hypertension; restenosis; arterial inflammation;			
KW	therapy; diagnosis; drug screening; marker.			
XX				
OS	Homo sapiens.			
XX				
PN	WO9730065-A1.			
XX				
PD	21-AUG-1997.			
XX				
PE	14-FEB-1997; 97WO-US02291.			
XX				
XX	13-FEB-1997; 97US-0799910.			
PR	16-FEB-1996; 96US-0011787.			
XX				
PA	(MILL-) MILLENNIUM PHARM INC.			
PI	Fald DA;			
XX				
DR	WPI; 1997-424966/39.			
DR	N-PSDB; AAT94469.			
XX				
PT	New genes differentially expressed in cardiovascular disease - used			
PT	for diagnosis; drug screening and treatment of cardiovascular			
PT	disease, e.g. atherosclerosis, restenosis, hypertension, etc			
XX				
PS	Example 7; Fig 3; 163pp; English.			
XX				
CC	This protein is encoded by the novel human fchd545 gene (see			
CC	AAT94469) that is down-regulated in endothelial cells subjected to			
CC	laminar shear stress. Shear stress is thought to be responsible			
CC	for the prevalence of atherosclerotic lesions in areas of unusual			
CC	circulatory flow. Novel fchd531, fchd540, fchd545, fchd602 and			
CC	fchd605 genes (see AAT94467-71) provide a fingerprint for the study			
CC	of cardiovascular diseases, including atherosclerosis,			
CC	ischaemia/reperfusion, hypertension, restenosis and arterial			
CC	inflammation. The fchd545 gene product is a transmembrane			
CC	protein that provides an excellent target for detection of			
CC	cardiovascular disease states in diagnostic systems, as well as in			
CC	monitoring the efficacy of compounds in clinical trials. Its			
CC	extracellular domains provide targets which allow the design of			
CC	efficient screening systems for identifying compounds that bind to			
CC	them. Such compounds can be useful in treating cardiovascular			
CC	diseases by modulating the activity of the transmembrane gene			
CC	product.			
XX				
SO	Sequence	283 AA;		
Query Match		0.7%; Score 7; DB 18; Length 283;		
Best Local Similarity		100.0%; Pred. No. 2,9e+02;		
Matches	7; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	782 KLAEGIK 788			
Db	90 klaeglk 96			
RESULT	59			

AAW48908	standard; Protein; 283 AA.
AAW48908	
23-SEP-1998	(first entry)
Human high voltage-dependent anion channel protein..	
HACH; human high voltage-dependent anion channel; genomic mapping; drug screening; proliferation disease; rheumatoid arthritis; tumour; immuno-diagnosis; hypothalamus CDNA library.	
Homo sapiens.	
Key	Location/Qualifiers
Domain	140..165
	/note="Membrane spanning domain"
US5780235-A.	
14-JUL-1998.	
04-OCT-1996;	96US-0726227.
04-OCT-1996;	96US-0726227.
(INCY-) INCYTE PHARM INC.	
Bandman O, Hillman JL;	
WPI: 1998-413045/35.	
N-PSDB: AAV32579.	
New high voltage dependent anion channel protein and related nucleic acid - vectors and transformed cells, useful for diagnosis and treatment of tumours and other proliferative diseases	
Claim 1; Fig 1A-1B; 25pp; English.	
The present sequence represents the Human high voltage-dependent anion channel (HACH) protein encoded by the HACH cDNA which was isolated from a hypothalamus CDNA library. Cells transformed with HACH cDNA can be used to produce recombinant HACH protein. HACH cDNA or its fragments, are claimed to be useful for detecting and/or quantifying HACH gene expression (for diagnosis or monitoring), as probes and primers for detecting genomic sequences encoding HACH or related proteins. They are also claimed to be useful in drug screening and genomic mapping. HACH protein or its activity is claimed to be useful for inhibiting growth of tumours and for treating other cell proliferation diseases, e.g. rheumatoid arthritis. HACH protein and its fragments are also claimed to be useful for screening binding agents for the protein, potential therapeutic agents, and to raise antibodies. Antibodies can be useful for diagnosing or monitoring HACH-related disorders, also therapeutically, in competitive drug screens, and for affinity purification of the HACH protein from natural sources.	
Sequence	283 AA:
Query Match	0.7%; Score 7; DB 19; Length 283;
Best Local Similarity	100.0%; Pred. No. 2.9e+02;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	782 KLAEGLK 788
Db	90 klaeglk 96
RESULT	60
AAV07222	
AAV07222	standard; Protein; 283 AA.

AC AAY07222;
XX
DT 16-JUL-1999 (first entry)
XX
DE Voltage-dependent anion channel CBMAD07 protein sequence.
XX
KW Human; voltage-dependent anion channel; CBMAD07; antibody; antagonist;
KM cancer; spontaneous abortion; infertility.
XX
OS Homo sapiens.
XX
PN WO921990-A1.
XX
PD 06-MAY-1999.
XX
PF 29-OCT-1997; 97WO-CN00118.
XX
PR 29-OCT-1997; 97WO-CN00118.
XX
PA (UYSH-) UNIV SHANGHAI SECOND MEDICAL.
XX
PI Wang Y, Zhang Q;
XX
DR WPI; 1999-303016/25.
DR N-PSDB; AAX57719.
XX
PT CBMAD07, a human voltage-dependent anion channel protein, useful in
PT the treatment and diagnosis of microsomal and neurological disorders
XX
PS Claim 11: Page 8; 31pp; English.
XX
CC This sequence represent a novel human voltage-dependent anion channel
CC designated CBMAD07. The protein, antibodies and (antagonists to it can
CC be used for treating, e.g. cancer, spontaneous abortion and infertility.
XX
SQ Sequence 283 AA;

Query Match 0.7%; Score 7; DB 20; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 782 KLAEGLK 788
DB 90 klaeglk 96
|||||

RESULT 61
AAY45015
ID AAY45015 standard; Protein; 283 AA.
XX
AC AAY45015;
XX
DT 31-MAY-2000 (first entry)
XX
DE Protein encoded by fchd545 gene.
XX
KW fchd545 gene; human; cardiovascular disease; oncogenic disorder;
KM diabetic retinopathy; fibroproliferative disorder; atherosclerosis;
KW TGF-beta signalling pathway; TGF; Transforming growth factor;
KM pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
KM vascularisation; cytostatic; antidiabetic; ophthalmological.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Region 48..63
FT /label= Antigenic_fragment
FT Region 107..121
FT /label= Antigenic_fragment
XX
PN WO200006206-A1.
XX

PD 10-FEB-2000.
XX
PF 30-JUL-1999; 99WO-US17394.
XX
PR 30-JUL-1998; 98US-0126640.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Falb DA;
XX
DR WPI; 2000-205414/18.
DR N-PSDB; AAZ50709.
XX
PT Identifying substances for ameliorating symptoms of fibroproliferative
PT diseases or oncogenic related disorders -
XX
PS Example; Fig 3; 214pp; English.
XX
CC The patent discloses methods for the treatment and diagnosis of
CC cardiovascular diseases by novel human genes which are differentially
CC expressed in different cardiovascular disease states. Compositions which
CC can modify TGF-beta signalling pathway are identified by screening.
CC These are used therapeutically to treat fibroproliferative and oncogenic
CC disorders, especially TGF (Transforming growth factor)-beta related
CC disorders, including diabetic retinopathy, atherosclerosis, pancreatic
CC cancer, angiogenesis, inflammation, fibrosis, tumour growth and
CC vascularisation. The present sequence is the protein product of fchd545
CC gene which is down-regulated in endothelial cells subjected to shear
CC stress can be used to design cardiovascular disease treatment
CC strategies. Depending on whether the down-regulation has a pathogenic or
CC protective effect treatment methods can be designed to increase or
CC decrease the activity of the protein product of the gene.
XX
SQ Sequence 283 AA;

Query Match 0.7%; Score 7; DB 21; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 782 KLAEGLK 788
DB 90 klaeglk 96
|||||

RESULT 62
AAB08519
ID AAB08519 standard; Protein; 287 AA.
XX
AC AAB08519;
XX
DT 20-DEC-2000 (first entry)
XX
DE Protein encoded by haemoglobin-response gene HBR2.
XX
KW Haemoglobin-response gene; HBR1; HBR2; HBR3; haemoglobin; adhesion;
KM disseminated infection; blastocyst; fibronectin.
XX
OS Candida albicans.
XX
PN WO200050601-A2.
XX
PD 31-AUG-2000.
XX
PF 18-JAN-2000; 2000WO-US01184.
XX
PR 26-FEB-1999; 99US-0258634.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Roberts DD, Van S;
XX
DR WPI; 2000-543913/49.
XX

XX Detecting disseminated *Candida albicans* infections using
 PT hemoglobin-response genes and proteins, probes and antibodies derived
 PT from them -
 XX
 XX
 PS Claim 3; Page 50-51; 54pp; English.
 CC
 CC The present sequence is encoded by a haemoglobin-response gene HBR2 from
 CC *Candida albicans*. The specification also describes haemoglobin-response
 CC genes HBR1 and HBR3. The expression of these genes is specifically
 CC induced when the organism is exposed to haemoglobin during disseminated
 CC infections. Haemoglobin induces increased adhesion of *C. albicans*
 CC blastoconidia to fibronectin. HBR1, HBR2 and HBR3 nucleic acid probes,
 CC proteins and antibodies are used for the diagnosis of disseminated
 CC *C. albicans* infections.
 CC
 SO Sequence 287 AA:
 OY
 Db 293 LTAEGVD 299
 |||||
 115 ltaevgd 121
 Query Match 0.7%; Score 7; DB 21; Length 287;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 63
 AAG72401
 ID AAG72401 standard; Protein: 295 AA.
 XX
 XX AAG72401;
 DT 30-JUL-2001 (first entry)
 XX
 XX Human OR-like polypeptide query sequence, SEQ ID NO: 2082.
 DE Human OR-like polypeptide query sequence, SEQ ID NO: 2082.
 XX
 XX Human; olfactory receptor; OR; primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation.
 XX
 OS *Homo sapiens*.
 XX
 XX WO200127158-A2.
 PN 19-APR-2001.
 PD
 XX
 XX 06-OCT-2000; 2000WO-US27582.
 PE
 XX 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.
 XX
 XX (DIGI-) DIGISENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 PA
 XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
 PI WPI; 2001-290713/30.
 DR
 XX
 PT New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 XX Example 6; Page 1396-1397; 1857pp; English.
 PS
 CC The present sequence is a polypeptide encoded by one of 344 newly mined
 CC human genes. It was used as a query sequence in a database search of
 CC olfactory receptor (OR)-like sequences. The invention relates to isolated
 CC polynucleotides encoding polypeptides involved in olfactory sensation.
 CC The polynucleotides can be used in screening for olfactory agonists and
 CC antagonists. The methods allow for the determination of primary scents
 CC and the identification of the odour receptors used to detect these
 CC primary scents. The methods also enable determination of secondary scents

CC and the identification of combinations of odour receptors that are
 CC involved in detecting such secondary scents. This enables the
 CC construction of a scent representation (also called a scent fingerprint
 CC or scent profile), which may be used to re-create and edit scents.
 CC Libraries of olfactory receptors are useful for determining the
 CC interaction pattern of a composition with the receptors, and can be
 CC used for determining differences in the olfactory faculties of different
 CC individuals.
 CC
 SO Sequence 295 AA:
 OY
 Db 680 LGRIPSS 686
 |||||
 286 lgrliss 292
 Query Match 0.7%; Score 7; DB 22; Length 295;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 64
 AAG23178
 ID AAG23178 standard; Protein: 297 AA.
 XX
 XX AAG23178;
 AC
 XX
 XX 17-OCT-2000 (first entry)
 DT
 XX
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 26387.
 DE
 XX
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 KW
 XX
 OS Arabidopsis thaliana.
 XX
 XX EP1033405-A2.
 PN 06-SEP-2000.
 PD
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 PF
 XX 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148655.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 0.7%; Score 7; DB 21; Length 297;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 759 LGNLRNL 765
Db 236 lgnlkn1 242
RESULT 65
AAG16591
ID AAG16591 standard; Protein; 298 AA.
XX AAG16591;
AC AAG16591;
XX 17-OCT-2000 (first entry)
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17295.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EP103405-A2.
PM 06-SEP-2000.
XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132867.
PR 11-MAY-1999; 99US-0132868.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144881.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.7%; Score 7; DB 21; Length 298;
Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 456 SSILTSH 462
Db 98 ssiltsn 104

RESULT 66

AAB95133

ID AAB95133 standard; Protein; 300 AA.

XX

XX

AC

XX

DT

XX

DE

XX

KM

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

XX

PR

XX

PR

XX

PR

XX

PI

XX

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 300 AA;

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Query Match 0.7%; Score 7; DB 22; Length 300;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSL 706
DB 104 VAGSLSL 110

RESULT 67

AAM35831

ID AAM35831 standard; Protein; 306 AA.

XX AAM35831;

XX 27-FEB-1998 (first entry)

XX Human G protein conjugate type receptor.

KW Human; G protein; guanine nucleotide binding protein; gene therapy;
KW receptor ligand; genetic diagnosis.

XX Homo sapiens.

XX JP09238686-A.

XX 16-SEP-1997.

XX 07-MAR-1996; 96JP-0050678.

XX 07-MAR-1996; 96JP-0050678.

XX (TAKE) TAKEDA CHEM IND LTD.

XX WPI; 1997-506555/47.

XX N-PSDB; AAT94894.

XX Novel G protein conjugate receptor - used for identifying receptor

XX ligands which may potentially be useful in therapeutic drugs

XX Claim 1; Page 23-24; 31pp; Japanese.

XX The present sequence represents a G protein (guanine nucleotide-binding

XX protein) conjugate-type receptor protein. The G protein and its

XX encoding DNA are used in the development of a receptor-binding assay

XX system for screening for candidate drugs. The DNA and fragments of

XX it may also be used as primers or probes for genetic diagnosis, and

XX in gene therapy. The elucidation of the structure and properties of

XX the G protein conjugate-type receptor is expected to lead to the

XX development of unique drugs acting on its system.

XX Sequence 306 AA;

XX

XX

XX

XX

XX

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 16-JUN-1999; 99US-0139453.

XX 17-JUN-1999; 99US-0139492.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139457.

XX 18-JUN-1999; 99US-0139458.

XX 18-JUN-1999; 99US-0139459.

XX 18-JUN-1999; 99US-0139460.

XX 18-JUN-1999; 99US-0139461.

XX 18-JUN-1999; 99US-0139462.

XX 18-JUN-1999; 99US-0139463.

XX 18-JUN-1999; 99US-0139750.

XX 18-JUN-1999; 99US-0139763.

XX 21-JUN-1999; 99US-0139817.

XX 22-JUN-1999; 99US-0139899.

XX 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142820.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143342.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 28-JUL-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146389.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.

PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 18-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160880.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 0.7%; Score 7; DB 21; Length 306;
Best Local Similarity 100.0%; Pred. No. 3; 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 704 LSLVLT 710
|||||
Db 300 LSLVLT 306

RESULT 69
AACG38088 standard; Protein; 306 AA.
XX
AC AAG38088;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46936.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.

XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0122180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-01300449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144321.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.7%; Score 7; DB 21; Length 306;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 704 LSLVLT 710
DB 300 LSLVLT 306

RESULT 70
AAB68564
ID AAB68564 standard; Protein; 307 AA.

AC AAB68564;

DT 19-APR-2001 (first entry)

DE Human GTP-binding associated protein #64.

XX Human; guanosine triphosphate binding associated protein; GTP; GBAP;
KW Inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;
KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
KW osteoporosis; psoriasis.

XX Homo sapiens.

XX WO200105970-A2.

XX 25-JAN-2001.

PD 19-JUL-2000; 2000MO-US19698.

PR 19-JUL-1999; 99US-0144595.

PR 23-AUG-1999; 99US-0150460.
PR 15-OCT-1999; 99US-0159849.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;
PI Reddy R, Yang J, Baughn MR, Lu DM, Azimzai Y, Patterson C;
XX
DR MPI: 2001-091972/10.
DR N-PSDB; AAF58364.
XX
PT New guanosine triphosphate-binding associated proteins (GBAP) and their
PT encoding nucleic acids, useful for treating and/or diagnosing diseases
PT associated with GBAP expression, such as cancer, diabetes and asthma -
XX
PS Claim 1; Pages 184-185; 233pp; English.
XX
CC The present invention relates to novel human guanosine triphosphate
CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their
CC coding sequences (AAF58301-AAF58366). The proteins and coding sequences
CC of the present invention are useful for treating a variety of disorders
CC including inflammation, AIDS, Addison's disease, anaemia,
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC psoriasis.
XX
SQ Sequence 307 AA:

Query Match 0.7%; Score 7; DB 22; Length 307;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 ADDIKDL 108
DB 174 adqldkl 180

RESULT 71

AA616590
ID AAG16590 standard; Protein; 316 AA.

AC AAG16590;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 17294.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP103405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0128845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132040.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138547.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149932.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.


```

PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

```

```

Query Match          0.7%; Score 7; DB 21; Length 316;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 456 SSLLTSH 462
DB 116 SSLLTSH 122

```

RESULT 72

```

AAM20728
ID AAM20728 standard; protein: 319 AA.

```

```

AC AAM20728;

```

```

DT 15-JUL-1997 (first entry)

```

```

DE H. pylori cytoplasmic protein, 06cel0515orf4.

```

```

KW Vaccine; prevention; treatment; infection; identification;
binding compound; bacterium; life cycle; activator; bacteria;
inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW cytoplasmic; mRNA; translation; ribosome; biogenesis.

```

```

OS Helicobacter pylori.

```

```

PN W09640893-A1.

```

```

PD 19-DEC-1996.

```

```

PF 06-JUN-1996; 96WO-US09122.

```

```

PR 01-APR-1996; 96US-0630405.

```

```

PR 07-JUN-1995; 95US-0487032.

```

```

PA (ASTR ) ASTRA AB.

```

```

PI Berglindh OT, Smith D, Mellgaard BL;

```

```

DR WPI; 1997-052306/05.

```

```

DR N-PSDB; AAT67981.

```

```

PT Helicobacter pylori nucleic acid sequences and related
polypeptide(s) - useful for vaccines to treat or prevent H. pylori
infection, and to detect Helicobacter

```

```

PS Claim 61; Pages 1143-1144; 1481pp; English.

```

```

CC The present sequence is a Helicobacter pylori cytoplasmic protein
involved in mRNA translation and ribosome biogenesis.
CC The protein may be used in a vaccine to prevent or treat

```

```

CC H. pylori infection or to identify H. pylori polypeptide binding
CC compounds, useful as potential H. pylori life cycle activators or
CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
CC determined from overlapping contigs generated by mechanically
CC shearing the bacterial DNA. The sequences were analysed for ORF of
CC at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from
CC various ORF were analysed for significant homology to other known
CC or exported membrane proteins. Having identified and determined
CC the sequences of interest, particular regions can be isolated from
CC H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.

```

```

SQ Sequence 319 AA;

```

```

Query Match          0.7%; Score 7; DB 18; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 67 NLFKSL 73
DB 232 NLFKSL 238

```

RESULT 73

```

AAR70153
ID AAR70153 standard; Protein: 320 AA.

```

```

AC AAR70153;

```

```

DT 14-FEB-1996 (first entry)

```

```

DE Streptococcus pneumoniae strain SPRU42 Exp2.

```

```

KW Exp2; export protein; pbp1a; p1pA; exp1; exp3; pad1;
virulence determinant; perasease like protein;

```

```

KW penicillin binding protein 1A; pyruvate oxidase; regulatory element;
KW acellular vaccine; antibody.

```

```

OS Streptococcus pneumoniae.

```

```

PN W09506732-A2.

```

```

PD 09-MAR-1995.

```

```

PF 01-SEP-1994; 94WO-US09942.

```

```

PR 01-SEP-1993; 93US-0116541.

```

```

PR 18-MAY-1994; 94US-0245511.

```

```

PA (UYRQ ) UNIV ROCKEFELLER.

```

```

PI Masure HR, Pearce BJ, Tuomanen E;

```

```

DR WPI; 1995-115448/15.

```

```

DR N-PSDB; AA083241.

```

```

PT Novel gene fragments encoding specific bacterial exported proteins
- specifically of S. pneumoniae, useful as vaccines

```

```

PS Claim 35; Page 88-9; 168pp; English.

```

```

CC This sequence represents exp2. The DNA encoding this sequence is
CC identical to that for pona which encodes penicillin-binding protein 1A
CC (PBP1a). This sequence is involved in adhesion of bacteria to target
CC cells. This sequence is an exported protein of S. pneumoniae. Export
CC proteins are the proteins in pathogenic bacteria that are virulence
CC determinants. Other export proteins include p1pA (see AAR70152), exp1,
CC exp3, and pad1 (encoded by the sequence shown in AA083259). This
CC sequence can be inserted into an expression vector (preferably a
CC bacterial expression vector) to provide for high levels of expression of

```

CC the protein. The protein can then be used in the production of an
CC acellular vaccine. These vaccines are used to provide protection from
CC Gram positive bacterial infection. Antibodies against export proteins
CC can be used for diagnosis of infection and in passive immune therapy.

SO Sequence 320 AA:

Query Match 0.7%; Score 7; DB 16; Length 320;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 924 IRLGAF 930
DB 55 IRLGAF 61

RESULT 74

ID AAG23177 standard; Protein: 324 AA.

XX AAG23177;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 26386.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.

PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142977.
PR 12-JUL-1999; 990S-0143297.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.7%; Score 7; DB 21; Length 324;
Best Local Similarity 100.0%; Pred No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 759 LGNKKNL 765
Db 263 LGNKKNL 269
RESULT 75
AAG23176
ID AAG23176 standard; Protein; 329 AA.
XX
AC AAG23176;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 26385.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999;	99US-0138647.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0138119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139859.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 30-AUG-1999;	99US-0151308.
PR 23-JUN-1999;	99US-0140354.	PR 31-AUG-1999;	99US-0151438.
PR 24-JUN-1999;	99US-0140695.	PR 01-SEP-1999;	99US-0151930.
PR 28-JUN-1999;	99US-0140823.	PR 07-SEP-1999;	99US-0152363.
PR 29-JUN-1999;	99US-0140991.	PR 10-SEP-1999;	99US-0153070.
PR 30-JUN-1999;	99US-0141287.	PR 13-SEP-1999;	99US-0153758.
PR 01-JUL-1999;	99US-0141842.	PR 13-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0142154.	PR 16-SEP-1999;	99US-0154039.
PR 02-JUL-1999;	99US-0142055.	PR 20-SEP-1999;	99US-0154779.
PR 06-JUL-1999;	99US-0142390.	PR 22-SEP-1999;	99US-0155139.
PR 08-JUL-1999;	99US-0142803.	PR 23-SEP-1999;	99US-0155486.
PR 09-JUL-1999;	99US-0142920.	PR 24-SEP-1999;	99US-0155659.
PR 12-JUL-1999;	99US-0142977.	PR 28-SEP-1999;	99US-0156458.
PR 13-JUL-1999;	99US-0143542.	PR 29-SEP-1999;	99US-0156596.
PR 14-JUL-1999;	99US-0143624.	PR 04-OCT-1999;	99US-0157117.
PR 15-JUL-1999;	99US-0144005.	PR 05-OCT-1999;	99US-0157753.
PR 16-JUL-1999;	99US-0144085.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144086.	PR 07-OCT-1999;	99US-0158029.
PR 19-JUL-1999;	99US-0144335.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144331.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144633.	PR 14-OCT-1999;	99US-0159637.
PR 21-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 21-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 22-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-01455913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 28-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146388.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161920.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147204.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147302.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		

Query Match 0.7%; Score 7; DB 21; Length 329;
Best local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 759 LGNKNL 765
|||||||

Tue Mar 26 11:36:04 2002

4 . 3.
Db 268 1gn1kn1 274

Search completed: March 25, 2002, 11:00:46
Job time: 241 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 11:01:11 ; Search time 35.05 Seconds
(without alignments)
4273.406 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 1024
Sequence: 1 MNFKIDNSRALIQRMGMVTI.....WQFDDDDSLVTGAFKLVTA 1024

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

1: SPRENBL_17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.todent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	346	33.8	439	4 Q9NPP4	Q9NPP4 homo sapien
2	9	0.9	1021	0 Q25200	Q25200 helicobacte
3	9	0.9	1120	10 Q9LP24	Q9LP24 arabidopsis
4	8	0.8	111	3 Q9URT8	Q9URT8 schizosacch
5	8	0.8	221	10 Q96477	Q96477 lycopersico
6	8	0.8	280	10 Q9LM24	Q9LM24 arabidopsis
7	8	0.8	350	5 P91506	P91506 caenorhabdi
8	8	0.8	356	2 Q9PLM5	Q9PLM5 campylobact
9	8	0.8	360	10 Q9LPV7	Q9LPV7 arabidopsis
10	8	0.8	434	5 Q9V7S7	Q9V7S7 drosophila
11	8	0.8	440	4 Q9PLX9	Q9PLX9 homo sapien
12	8	0.8	440	6 Q9N2A7	Q9N2A7 pan troglod
13	8	0.8	440	6 Q9N2A6	Q9N2A6 gorilla gor
14	8	0.8	440	6 Q9N2A5	Q9N2A5 pongo pygma
15	8	0.8	493	2 Q08355	Q08355 pseudomonas
16	8	0.8	680	10 Q23225	Q23225 arabidopsis
17	8	0.8	902	5 Q17576	Q17576 caenorhabdi
18	8	0.8	928	5 Q9BLX1	Q9BLX1 caenorhabdi
19	8	0.8	982	10 Q9FWMO	Q9FWMO oryza sativ

20	8	0.8	1040	4 Q9HC29	Q9HC29 homo sapien
21	8	0.8	1112	10 Q41397	Q41397 lycopersico
22	8	0.8	1112	10 Q41398	Q41398 lycopersico
23	8	0.8	1141	10 Q9LHP4	Q9LHP4 arabidopsis
24	8	0.8	1200	13 Q91054	Q91054 heterodontu
25	8	0.7	78	6 Q9GL29	Q9GL29 bos taurus
26	7	0.7	95	4 Q9Y6U4	Q9Y6U4 homo sapien
27	7	0.7	97	3 Q14396	Q14396 schizosacch
28	7	0.7	103	1 Q9YD13	Q9YD13 aeropyrum p
29	7	0.7	107	3 Q05413	Q05413 saccharomyc
30	7	0.7	110	13 Q91AB7	Q91AB7 brachydanio
31	7	0.7	111	2 Q9PEQ1	Q9PEQ1 xylella fas
32	7	0.7	120	13 Q91AB8	Q91AB8 brachydanio
33	7	0.7	122	5 Q9GXU3	Q9GXU3 leishmania
34	7	0.7	122	10 Q9ANR6	Q9ANR6 pisum sativ
35	7	0.7	133	2 Q9LAB5	Q9LAB5 streptomyce
36	7	0.7	138	1 Q9YBU7	Q9YBU7 aeropyrum p
37	7	0.7	140	8 Q9G4E5	Q9G4E5 fasciola gi
38	7	0.7	142	2 Q9RHM6	Q9RHM6 porphyromon
39	7	0.7	146	2 Q9HYM6	Q9HYM6 pseudomonas
40	7	0.7	146	10 Q91ZNI	Q91ZNI arabidopsis
41	7	0.7	157	2 Q9PHF7	Q9PHF7 xylella fas
42	7	0.7	158	6 Q19028	Q19028 macaca radi
43	7	0.7	161	2 Q9K5D6	Q9K5D6 campylobact
44	7	0.7	164	2 Q9KHE3	Q9KHE3 anabaena sp
45	7	0.7	170	4 Q14843	Q14843 homo sapien
46	7	0.7	170	6 Q28710	Q28710 oryctolagus
47	7	0.7	171	2 Q53048	Q53048 rhodococcus
48	7	0.7	176	8 Q63507	Q63507 echinostoma
49	7	0.7	176	8 Q63508	Q63508 echinostoma
50	7	0.7	176	8 Q63509	Q63509 echinostoma
51	7	0.7	178	1 Q28975	Q28975 archaeoglob
52	7	0.7	179	5 Q9GXM7	Q9GXM7 leishmania
53	7	0.7	181	2 Q53573	Q53573 synechococc
54	7	0.7	188	11 Q04365	Q04365 mus musculu
55	7	0.7	210	2 Q9KEN8	Q9KEN8 bacillus ha
56	7	0.7	218	10 Q9FPU5	Q9FPU5 arabidopsis
57	7	0.7	221	2 Q9PM21	Q9PM21 campylobact
58	7	0.7	222	2 Q83565	Q83565 treponema p
59	7	0.7	222	5 Q9VRES	Q9VRES drosophila
60	7	0.7	226	12 Q9YMU9	Q9YMU9 lymantria d
61	7	0.7	227	2 Q9AD75	Q9AD75 streptomyce
62	7	0.7	229	4 Q9ULR9	Q9ULR9 homo sapien
63	7	0.7	231	4 Q95372	Q95372 homo sapien
64	7	0.7	231	11 Q9QYL8	Q9QYL8 rattus norv
65	7	0.7	231	11 Q9WTL7	Q9WTL7 mus musculu
66	7	0.7	233	2 Q9Z8V6	Q9Z8V6 chlamydia p
67	7	0.7	235	2 Q9RWN3	Q9RWN3 delnoccocus
68	7	0.7	236	5 Q26543	Q26543 schistosoma
69	7	0.7	237	2 Q46699	Q46699 e.pkm101.co
70	7	0.7	237	2 Q916G5	Q916G5 escherichia
71	7	0.7	240	2 Q25044	Q25044 helicobacte
72	7	0.7	242	2 P96389	P96389 mycobacteri
73	7	0.7	244	11 Q9CWL7	Q9CWL7 mus musculu
74	7	0.7	246	2 Q91LOJ9	Q91LOJ9 streptomyce
75	7	0.7	256	2 Q9K7R4	Q9K7R4 bacillus ha
76	7	0.7	259	11 Q63968	Q63968 mus sp. cyt
77	7	0.7	259	12 Q41131	Q41131 paramecium
78	7	0.7	264	1 Q27788	Q27788 methanobact
79	7	0.7	264	1 Q9YEX5	Q9YEX5 aeropyrum p
80	7	0.7	264	2 Q9XAA7	Q9XAA7 streptomyce
81	7	0.7	265	4 Q9Y3A0	Q9Y3A0 homo sapien
82	7	0.7	266	2 Q9WY65	Q9WY65 thermotoga
83	7	0.7	266	5 Q18862	Q18862 caenorhabdi
84	7	0.7	268	2 Q52433	Q52433 pseudomonas
85	7	0.7	271	2 Q47797	Q47797 enterococcu
86	7	0.7	276	2 P70970	P70970 bacillus su
87	7	0.7	278	2 Q99Y44	Q99Y44 streptococc
88	7	0.7	281	2 Q9F1G4	Q9F1G4 enterococcu
89	7	0.7	283	2 Q9FPL7	Q9FPL7 xylella fas
90	7	0.7	283	6 Q9W214	Q9W214 sus scrofa
91	7	0.7	283	11 Q9J131	Q9J131 rattus norv
92	7	0.7	283	11 Q9ESR2	Q9ESR2 rattus norv

ID	Accession	Score	Length	DB	Score	DB	Length	DB
93	7	0.7	284	2	0.9KUR8			
94	7	0.7	284	10	P93506			
95	7	0.7	288	2	O9CIS8			
96	7	0.7	292	2	O9KMO6			
97	7	0.7	294	2	O9A7V7			
98	7	0.7	296	2	O9X772			
99	7	0.7	297	9	O37909			
100	7	0.7	298	2	O9RHNS			

ALIGNMENTS

RESULT 1

Q9NPP4 PRELIMINARY; PRT: 439 AA.

AC Q9NPP4; 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DE HYPOTHEICAL 49.3 KDA PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Pluvinet R., Estvill X., Escarceller M., Sumoy L.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [12]

RP SEQUENCE FROM N.A.

RA Aulfray C., Ansozge W., Ballabio A., Estvill X., Gibson K., Lehnach H., Poultka A., Lundberg J.;

RT "The European IMAGE consortium for Integrated Molecular analysis of human gene transcripts.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL389934; CAB97523.1; -

KW Hypothetical protein.

FT NON_TER

SO SEQUENCE 439 AA; 49286 MW; 7EFC9C8F0B7359A CRC64;

Query Match 33.8%; Score 346; DB 4; Length 439;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 679 YLGRFSSATSLRLQIKRCAGVAGSLVLTCKNIYSLMWEASPLTIEDERHITSVNL 738

DB 94 YLGRFSSATSLRLQIKRCAGVAGSLVLTCKNIYSLMWEASPLTIEDERHITSVNL 153

QY 739 KTLSTHDLQNRPLPGGLTDSLGKLNKLTLMNDIKMNEEDAIKLAEGIKLNKMKCFHL 798

DB 154 KTLSTHDLQNRPLPGGLTDSLGKLNKLTLMNDIKMNEEDAIKLAEGIKLNKMKCFHL 213

QY 799 THLSDIGGMIVYVSSSEPDLEIOLVSCISANAVKILAQNLHNLVYKSTIIDSEN 858

DB 214 THLSDIGGMIVYVSSSEPDLEIOLVSCISANAVKILAQNLHNLVYKSTIIDSEN 273

QY 859 YLEKDGNALEHLLDRMVNLEOLATMLPMGCDVGSLSLKLHEEVPOLYKLGKMKR 918

DB 274 YLEKDGNALEHLLDRMVNLEOLATMLPMGCDVGSLSLKLHEEVPOLYKLGKMKR 333

QY 919 LDTDEIRILGAFEGKNPLKNEQOILAGNRVSSDGLAFMGVFNMLKOLVFEFSTKEFL 978

DB 334 LDTDEIRILGAFEGKNPLKNEQOILAGNRVSSDGLAFMGVFNMLKOLVFEFSTKEFL 393

QY 979 PDPALVRLSOVLSTFLQEARLYGNQFDDDLISVITGAFRLTYA 1024

DB 394 PDPALVRLSOVLSTFLQEARLYGNQFDDDLISVITGAFRLTYA 439

RESULT 2

025200

ID 025200 PRELIMINARY; PRT: 1021 AA.

AC 025200;

DT 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)

DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE HYPOTHEICAL 120.1 KDA PROTEIN.

GN HP0453.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.

OX NCBI_TaxID=210;

RN [11]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kersavage A.R., Klenk H.-P., Gill S., Dougherty B.A., Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A., Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S., Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A., McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K., Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M., Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E., Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M., Venter J.C.;

RA "The complete genome sequence of the gastric pathogen Helicobacter pylori.";

RT Nature 388:539-547(1997).

RL Nature 388:539-547(1997).

DR EMBL: AE000560; AAD07519.1; -

DR TIGR: HP0453; -

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 1021 AA; 120112 MW; D688694561694469 CRC64;

Query Match 0.9%; Score 9; DB 2; Length 1021;

Best Local Similarity 100.0%; Pred. No. 9.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 735 VTNLKTLST 743

DB 879 VTNLKTLST 887

RESULT 3

Q9LP24 PRELIMINARY; PRT: 1120 AA.

AC Q9LP24;

DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE F14D7.1 PROTEIN.

GN F14D7.1.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [11]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Liu S.X., Chan A., Sakano H., Yu G., Lee J.M., Lenz C., Pham P., Tortum M., Chin C., Chou J., Choi E., Chung M., Gonzalez A., Huang B., Liu A., Vaysberg M., Altati H., Brooks S., Buehler E., Chao Q., Conn L., Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shin P., Southwick A., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A., Yu G.;

RA "The sequence of BAC F14D7 from Arabidopsis thaliana chromosome 1.";

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [12]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC211198; AAF79881.1; -
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_tyr.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00560; LRR; 23.
 DR Pfam: PF00659; pkinase; 1.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR; 22.
 DR SMART: SM00221; STYKC; 1.
 DR SMART: SM00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KM ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1120 AA; 124103 MW; DF5A5A0BA67B1357 CRC64;

Query Match 0.9%; Score 9; DB 10; Length 1120;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 SLGNLKNLT 766
 |||||
 DB 145 SLGNLKNLT 153

RESULT 4
 Q9URT8 PRELIMINARY; PRT; 111 AA.
 ID Q9URT8
 AC Q9URT8; 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE 60S RIBOSOMAL PROTEIN L34-B.
 GN SPC1322.15.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Lucas M., Galliard C., Lyne M., Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL: AL035259; CA22868.1; -
 DR InterPro: IPR001284; Ribosomal_L34E.
 DR Pfam: PF01199; Ribosomal_L34e; 1.
 DR PRINTS: PD005148; Ribosomal_L34E; 1.
 DR ProDom: PD005148; Ribosomal_L34E; 1.
 DR PROSITE: PS01145; RIBOSOMAL_L34E; 1.
 KM Ribosomal protein.
 SQ SEQUENCE 111 AA; 12800 MW; B0148F3079677F26 CRC64;

Query Match 0.8%; Score 8; DB 3; Length 111;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CLSANAVK 838

DB 79 CLSANAVK 86
 |||||

RESULT 5
 Q96477 PRELIMINARY; PRT; 221 AA.
 ID Q96477
 AC Q96477;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE LRR PROTEIN.
 GN LRR.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. VFN8; TISSUE-LEAF.
 RX MEDLINE-96367673; PubMed-8771787;
 RA Tornero P., Mayda E., Gomez M.D., Canas L., Conejero V., Vera P.;
 RT "Characterization of LRP, a leucine-rich repeat (LRR) protein from
 tomato plants that is processed during pathogenesis.";
 RL Plant J. 10:315-330(1996).
 DR EMBL: X95269; CA64565.1; -
 DR Mendel; 14339; Lyces; 2362; 14339.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 4.
 DR SMART: SM00370; LRR; 3.
 SQ SEQUENCE 221 AA; 24188 MW; 4C57C7EFFBF4058 CRC64;

Query Match 0.8%; Score 8; DB 10; Length 221;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 SLGNLKNLT 765
 |||||
 DB 140 SLGNLKNLT 147

RESULT 6
 Q9LM24 PRELIMINARY; PRT; 280 AA.
 ID Q9LM24
 AC Q9LM24; 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE T1002.23.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
 RA Kim C., Altafi H., Bel Q., Chin Q., Chlou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharshy N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC T10022 from chromosome
 I.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC069551; AAF78380.1; -
 SQ SEQUENCE 280 AA; 32809 MW; 774573C5F956FF7 CRC64;

```

Query Match          0.8%; Score 8; DB 10; Length 280;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 233 MAMLIKLR 240
    |||||||
Db 226 MAMLIKLR 233

RESULT 7
P91506 PRELIMINARY: PRT: 350 AA.
AC P91506;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILAR TO OTHER C. ELEGANS CHEMORECEPTOR.
GN T28A11.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Anscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightner J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Spaldon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Mohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA Rohlfing T.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U80027; AAC48131.1; -
DR InterPro: IPR000168; 7TM_nematode.
DR InterPro: IPR003002; 7TM_chemorecept_1.
DR Pfam: PF01461; 7tm_4; 1.
SQ SEQUENCE 350 AA; 39804 MW; B5B3A622DDFB253D CRC64;

Query Match          0.8%; Score 8; DB 5; Length 350;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 686 SATSLRLQ 693
    |||||||
Db 237 SATSLRLQ 244

RESULT 8
Q9PLW5 PRELIMINARY: PRT: 356 AA.
AC Q9PLW5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN CUI713.
GN Cui713.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCYC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jagers K., Kariyeh A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL, AL139079; CAB73699.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 356 AA; 40254 MW; 23DAEC477285E065 CRC64;

Query Match          0.8%; Score 8; DB 2; Length 356;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 836 AVKILACN 843
    |||||||
Db 184 AVKILACN 191

RESULT 9
Q9LP7 PRELIMINARY: PRT: 360 AA.
AC Q9LP7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE F15H18.23.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei B., Chin C., Chio J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thayerl A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F15H18 from chromosome
RT I."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA Shin P., Altafi H., Bei B., Chin C., Chio J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thayerl A.,
RA Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N., Theologis A.,
RA Ecker J.R.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chio J., Choi E.,

```

RA Conñ L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharisky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thayer A., Toriumi M., Vaysberg M., Yu C., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.,
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC013354; AAF25991.1;
 SQ SEQUENCE 360 AA; 41605 MW; 468AAAFCD8D2749E CRC64;

Query Match 0.8%; Score 8; DB 10; Length 360;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 233 MAMLUKLR 240
 Db 226 MAMLUKLR 233

RESULT 10
 ID 09V7S7 PRELIMINARY; PRT; 434 AA.
 AC 09V7S7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG5197 PROTEIN.
 GN CG5197
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bence P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck Y., Broxstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Eppanellista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spiedling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster";
 SC Science 287:2185-2195(2000).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL: AE003806; AAF57967.1;
 DR HSSP: P00763; IDPO.
 DR IYBase: F890034147; CG5197.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPE; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 434 AA; 47233 MW; 208F735989F06C81 CRC64;

Query Match 0.8%; Score 8; DB 5; Length 434;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 846 NLVKLSIL 853
 Db 183 NLVKLSIL 190

RESULT 11
 ID 09P1X9 PRELIMINARY; PRT; 440 AA.
 AC 09P1X9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MUSCARINIC ACETYLCHOLINE RECEPTOR M2 (FRAGMENT).
 GN CHRM2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kitano T., Kobayakawa H., Saitou N.;
 RT "Silver Project";
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 DR EMBL: AB041391; BAA94476.1;
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm.1; 1.
 DR PRINTS: PR00237; GPCR_RHODPSN.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_FL_2; 1.
 KW G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
 FT NON_TER
 SQ SEQUENCE 440 AA; 48853 MW; 12B0324E13D37DDF CRC64;

Query Match 0.8%; Score 8; DB 4; Length 440;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 700 VAGSISLV 707
 Db 3 VAGSISLV 10

RESULT 12
 ID 09N2A7 PRELIMINARY; PRT; 440 AA.
 AC 09N2A7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MUSCARINIC ACETYLCHOLINE RECEPTOR M2 (FRAGMENT).
GN CHRM2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHIM-220;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DE EMBL; AB041392; BAA94477.1; -;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCRHHODPSN.
DR PROSITE: PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
FT NON_TER
SQ SEQUENCE 440 AA; 48853 MW; 12B0324E13D37DDF CRC64;

Query Match 0.8%; Score 8; DB 6; Length 440;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSLV 707
|||||
DB 3 VAGSLSLV 10

RESULT 13
O9N2A6 PRELIMINARY; PRT; 440 AA.
AC O9N2A6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MUSCARINIC ACETYLCHOLINE RECEPTOR M2 (FRAGMENT).
GN CHRM2.
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GORILLA-UI;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DE EMBL; AB041393; BAA94478.1; -;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCRHHODPSN.
DR PROSITE: PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
FT NON_TER
SQ SEQUENCE 440 AA; 48853 MW; 12B0324E13D37DDF CRC64;

Query Match 0.8%; Score 8; DB 6; Length 440;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 700 VAGSLSLV 707
|||||

DB 3 VAGSLSLV 10

RESULT 14
O9N2A5 PRELIMINARY; PRT; 440 AA.
AC O9N2A5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MUSCARINIC ACETYLCHOLINE RECEPTOR M2 (FRAGMENT).
GN CHRM2.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORAN-UI;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DE EMBL; AB041394; BAA94479.1; -;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCRHHODPSN.
DR PROSITE: PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
FT NON_TER
SQ SEQUENCE 440 AA; 48869 MW; BDB2FEB3C1AC3B83E CRC64;

Query Match 0.8%; Score 8; DB 6; Length 440;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSLV 707
|||||
DB 3 VAGSLSLV 10

RESULT 15
O08355 PRELIMINARY; PRT; 493 AA.
AC O08355;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MANNITOL 2-DEHYDROGENASE (EC 1.1.1.67) (MDH).
GN MTL2.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND SEQUENCE OF 1-20.
RC STRAIN=DSM 50106;
RX MEDLINE=97236441; PubMed=9116029;
RA Bruenker P., Altenbuchner J., Kulbe K.D., Mattes R.;
RT "Cloning, nucleotide sequence and expression of a mannitol dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in Escherichia coli";
RL Biochim. Biophys. Acta 1351:157-167(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 50106;
RX MEDLINE=98121321; PubMed=9461423;
RA Bruenker P., Altenbuchner J., Mattes R.;
RT "Structure and function of the genes involved in mannitol, arabitol and glucitol utilization from Pseudomonas fluorescens DSM50106";

RL Gene 206:117-126(1998).
 CC -1- FUNCTION: A MANNITOL DEHYDROGENASE WITH A BROAD SUBSTRATE
 CC SPECIFICITY. SUBSTRATES INCLUDE MANNITOL, ARABITOL AND SORBITOL,
 CC THESE ARE OXIDIZED TO GIVE THE CORRESPONDING KETO SUGARS. THIS
 CC ENZYME WILL ALSO CATALYZE THE REDUCTION OF FRUCTOSE AND XYLULOSE.
 CC -1- CATALYTIC ACTIVITY: D-MANNITOL + NAD(+) = D-FRUCTOSE + NADH.
 CC -1- SUBUNIT: MONOMER.
 CC -1- INDUCTION: BY MANNITOL, ARABITOL AND SORBITOL. NOT INDUCED BY
 CC FRUCTOSE OR XYLULOSE.
 CC -1- MISCELLANEOUS: THE MTLD PROTEIN IS ENCODED BY THE MTL EFGKDYZ
 CC OPERON. THIS OPERON ENCODES PROTEINS FOR THE UPTAKE AND
 CC UTILIZATION OF MANNITOL, ARABITOL AND SORBITOL.
 CC -1- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
 DR EMBL: AF007800; AAC04472.1; -.
 DR InterPro: IPR000669; Mannitol_dh.
 DR Pfam: PF01232; Mannitol_dh.1.
 DR PRINTS: PR00084; MTLDHGNASE.
 DR PROSITE: PS00974; MANNITOL_DHCENASE.1.
 DR OXidoreductase; NAD.
 DR NP_BIND 29 40 NAD (BY SIMILARITY).
 SQ SEQUENCE 493 AA; 54497 MW; 7C12DFA43CEAA43 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 493;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 302 EDSQAALI 309
 |||||
 DB 109 EDSQAALI 116

RESULT 16
 ID 023225 PRELIMINARY; PRT; 680 AA.
 AC 023225;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOHETICAL 75.3 KDA PROTEIN.
 DE C7A10.810 OR ATG636550.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Terry N., Vos P., Heijnen L., Mewes H.W., Scheller C.,
 RA Chailatiz N.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z99708; CAB1638.1; -.
 DR EMBL: AL161589; CAB80321.1; -.
 DR InterPro: IPR003613; Ubox.
 DR SMART: SM00504; Ubox; 1.
 DR Hypothetical protein.
 KW SEQUENCE 680 AA; 75298 MW; 359CABCF842A2656 CRC64;

Query Match 0.8%; Score 8; DB 10; Length 680;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 894 GSLSLTK 901
 |||||
 DB 465 GSLSLTK 472

RESULT 17

017576
 ID 017576 PRELIMINARY; PRT; 902 AA.
 AC 017576;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE KIN-8 PROTEIN (RECEPTOR TYROSINE KINASE).
 GN KIN-8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Berks M.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; Pubmed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mortimore B.J.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RA Koga M., Takeuchi M., Tameishi T., Ohshima Y.;
 RT "Control of DAF-7 TGF expression and neuronal process development by a
 RT receptor tyrosine kinase KIN-8 in C. elegans.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL: Z35595; CAA84639.2; -.
 DR EMBL: Z47808; CAA84639.2; JOINED.
 DR EMBL: Z47808; CAB61003.1; -.
 DR EMBL: Z35595; CAB61003.1; JOINED.
 DR EMBL: AJ132947; CAC29085.1; -.
 DR HSSP: P08631; IAD5.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000024; Fz_domain.
 DR InterPro: IPR003598; Ig_L2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR000001; Kringle.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF01392; Fz; 1.
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF00051; kringle; 2.
 DR Pfam: PF00069; kinase; 1.
 DR PRINTS: PR00108; TYRKINASE.
 DR PRINTS: PR00109; TYRKINASE.
 DR SMART: SM00408; IGC2.1.
 DR SMART: SM00130; KR; 1.
 DR SMART: SM00219; TYKIC.1.
 DR PROSITE: PS00021; KRINGLE_1; FALSE_NEG.
 DR PROSITE: PS50070; KRINGLE_2; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; kinase; transferase; tyrosine-protein kinase.
 SQ SEQUENCE 902 AA; 101038 MW; ZA03D76D07C552B5 CRC64;

Query Match 0.8%; Score 8; DB 5; Length 902;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 713 NIYSLWE 720
 |||||
 DB 805 NIYSLWE 812

RESULT 18
 09BLY1 PRELIMINARY; PRT; 928 AA.
 AC 09BLY1;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE RECEPTOR TYROSINE KINASE.
 GN KIN-8
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2.
 RA Koga M., Takeuchi M., Tameishi T., Ohshima Y.;
 RT "Control of DAF-7 TGF expression and neuronal process development by a
 RT receptor tyrosine kinase KIN-8 in C. elegans."
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ132946; CAC29084.1; -
 KW kinase
 SQ SEQUENCE 928 AA; 103864 MW; F1388C9BCAB30D20 CRC64;

Query Match 0.8%; Score 8; DB 5; Length 928;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 713 NYSLME 720
 |||||
 DB 831 NYSLME 838

RESULT 19
 09FWM0 PRELIMINARY; PRT; 982 AA.
 AC 09FWM0;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE DISEASE RESISTANCE PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaeae; Oryza.
 OX NCBI_TaxID=4350;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Buell C.R., Yuan O., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
 RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBa0079L16 genomic sequence."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC026815; AAG21909.1; -
 OX NCBI_TaxID=4084;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CF 2;
 RA MEDLINE=96190812; PubMed=8608599;
 RA Dixon M.S., Jones D.A., Kedde J.S., Thomas C.M., Harrison K.,
 RA Jones J.D.G.;
 RT "The tomato Cf-2 disease resistance locus comprises two functional
 RT genes encoding leucine-rich repeat proteins."
 RL Cell 84:451-459(1996).
 DR EMBL: U42444; AAC15779.1; -
 DR Medel: 14333; Solp1.2357;14333.
 DR InterPro: IPR001611; LRR.
 DR SMART: SM00369; LRR_TYP: 9.
 DR SMART: SM00369; LRR_TYP: 9.
 SQ SEQUENCE 982 AA; 106421 MW; FADCF3EB228D8F1 CRC64;

Query Match 0.8%; Score 8; DB 10; Length 982;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 759 LGNLKNTL 766
 |||||
 DB 336 LGNLKNTL 343

RESULT 20
 09HC29 PRELIMINARY; PRT; 1040 AA.
 AC 09HC29;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NOD2 PROTEIN.
 GN NOD2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ogura Y., Inohara N., Benito A., Chen F.F., Yamaoka S., Nunez G.;
 RT "Nod2, a Nod1/Apa1 family member that is restricted to monocytes and
 RT activates NF-kappaB."
 RL J. Biol. Chem. 0:0-0(2001).
 DR EMBL: AF178930; AAG33677.1; -
 DR InterPro: IPR001687; ATP_GTP_A.
 DR InterPro: IPR001315; CARD.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003590; LRR_RNinh.
 DR Pfam: PF00560; LRR: 2.
 DR SMART: SM00114; CARD: 1.
 DR SMART: SM00370; LRR: 5.
 DR SMART: SM00368; LRR_RIT: 7.
 DR PROSITE: PS50209; CARD: 1.
 SQ SEQUENCE 1040 AA; 115282 MW; 0037592D96D7DDFF CRC64;

Query Match 0.8%; Score 8; DB 4; Length 1040;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLOR 181
 |||||
 DB 304 GKSTLLOR 311

RESULT 21
 041397 PRELIMINARY; PRT; 1112 AA.
 AC 041397;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CF-2-1.
 OS Lycopersicon pimpinellifolium (current tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4084;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CF 2;
 RA MEDLINE=96190812; PubMed=8608599;
 RA Dixon M.S., Jones D.A., Kedde J.S., Thomas C.M., Harrison K.,
 RA Jones J.D.G.;
 RT "The tomato Cf-2 disease resistance locus comprises two functional
 RT genes encoding leucine-rich repeat proteins."
 RL Cell 84:451-459(1996).
 DR EMBL: U42444; AAC15779.1; -
 DR Medel: 14333; Solp1.2357;14333.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.

DR Pfam: PF00560; LRR: 30.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR: 35.
 SQ SEQUENCE 1112 AA: 122188 MW: 60C8CE5278BF56E CRC64;

Query Match 0.8%; Score 8; DB 10; Length 1112;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 758 SLGNLKNL 765
 |||||||
 Db 330 SLGNLKNL 337

RESULT 22
 ID Q41398 PRELIMINARY; PRT; 1112 AA.
 AC Q41398;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE Cf-2.2.
 OS Lycopersicon pimpinellifolium (currant tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4084;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-CF 2;
 RC MEDLINE=66190812; PubMed=6608599;
 RA Dixon M.S., Jones D.A., Keddie J.S., Thomas C.M., Harrison K.,
 RA Jones J.D.G.,
 RT "The tomato Cf-2 disease resistance locus comprises two functional
 RT genes encoding leucine-rich repeat proteins."
 RL Cell 84:451-459(1996).
 DR EMBL: U42445; AAC15780.1;
 DR Mendel: 14334; Solp1;2357;14334.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR: 30.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR: 35.
 SQ SEQUENCE 1112 AA: 122190 MW: 111D20B296BA07A CRC64;

Query Match 0.8%; Score 8; DB 10; Length 1112;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 758 SLGNLKNL 765
 |||||||
 Db 330 SLGNLKNL 337

RESULT 23
 ID Q9LHP4 PRELIMINARY; PRT; 1141 AA.
 AC Q9LHP4;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE RECEPTOR PROTEIN KINASE-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eusteroideae; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-COLUMBIA;
 RC Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RP [12]
 RN SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RC PubMed=10907853;
 RA Nakamura Y.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: AP002037; BAB03091.1;
 DR EMBL: AB028621; BAB03091.1; JOINED.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR002290; Ser_thr_kin_actsite.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00560; LRR: 14.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR: 22.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 1141 AA: 124502 MW: 1C9CE94DADCT78B01 CRC64;

Query Match 0.8%; Score 8; DB 10; Length 1141;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 849 KLSILDLS 856
 |||||||
 Db 636 KLSILDLS 643

RESULT 24
 ID Q91054 PRELIMINARY; PRT; 1200 AA.
 AC Q91054;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CD45 HOMOLOG (EC 3.1.3.48).
 OS Heterodontus francisci (Horn shark).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontidae;
 OC Heterodontus.
 OC NCBI_TaxID=7792;
 OX [1]
 RN SEQUENCE FROM N.A.
 RA Okumura M., Matthews R.J., Robb B., Bork P., Thomas M.L.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U34750; AAB01087.1;
 DR HSP: P18052; IYFO.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR000387; Tyr_phosphatase.
 DR InterPro: IPR000242; Tyr_prot_phptase.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF00102; X-phosphatase; 2.
 DR PRINTS: PR00700; PRTYPHPTASE.
 DR SMART: SM00194; PTPc; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolase.
 SQ SEQUENCE 1200 AA: 135372 MW: EFC6B2B4DC02BC2 CRC64;

Query Match 0.8%; Score 8; DB 13; Length 1200;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0;

Oy 236 LKLRORV 243
| | | | | | | |
Db 729 LKLRORV 736

RESULT 25

O9GL29 PRELIMINARY; PRT; 78 AA.

AC O9GL29;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE VOLTAGE-DEPENDENT ANION CHANNEL 3 (FRAGMENT).
CN VDAC3.

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Konrad L., Asmarinah, Hirsch E., Hirsch K.D.;
RT "Expression of porins in the testis."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ294423; CAC14092.1; -.

FT NON_TER 1 78
SQ SEQUENCE 78 AA; 8679 MW; 389DBF04D92D50ED CRC64;

Query Match 0.7%; Score 7; DB 6; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Oy 782 KLAEGLK 788
| | | | | | | |
Db 46 KLAEGLK 52

RESULT 26

O9Y6U4 PRELIMINARY; PRT; 95 AA.

AC O9Y6U4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE CAMP RESPONSIVE ELEMENT BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;

RT "Toward a complete human genome sequence."
RL Genome Res. 8:1097-1108(1998).
RN [2]

RP SEQUENCE FROM N.A.
RA Kramer J., Sun H., Biewald T.;

RT "The sequence of Homo sapiens BAC clone RG491N20."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.
RA Waterston R.H.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Waterston R.H.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [5]
RP SEQUENCE FROM N.A.
RA Waterston R.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005105; AAD43180.1; -.
FT NON_TER 95 95
SQ SEQUENCE 95 AA; 10068 MW; D92EA815902CD66B CRC64;

Oy 894 GSLSL 900
| | | | | | | |
Db 20 GSLSL 26

RESULT 27

O14396 PRELIMINARY; PRT; 97 AA.

AC O14396;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE HYPOTHETICAL 10.3 KDA PROTEIN (FRAGMENT).
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetes; Schizosaccharomycetaceae;

OX NCBI_TaxID=4896;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Jang Y.J., Yoo H.S.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97397; AAB63869.1; -.

DR InterPro; IPR001284; Ribosomal_L34E.
DR Pfam; PF01199; Ribosomal_L34E; 1.
DR PRINTS; PR01250; RIBOSOMAL_L34.
DR ProDom; PD005148; Ribosomal_L34E; 1.
KW Hypothetical protein.
FT NON_TER 1 97
SQ SEQUENCE 97 AA; 10261 MW; CD75E5722FC7C6B7 CRC64;

Query Match 0.7%; Score 7; DB 3; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 831 CLSANAV 837
| | | | | | | |
Db 83 CLSANAV 89

RESULT 28

O9YD13 PRELIMINARY; PRT; 103 AA.

AC O9YD13;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE HYPOTHETICAL 10.7 KDA PROTEIN APE1099.
GN APE1099.

OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OX Aeropyrum.

OX NCBI_TaxID=56636;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=KL;
RX MEDLINE=99310339; PubMed=10382966;

RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Hino Y., Takahashi M., Sekine M., Baba S.-I., Aikai A., Kosugi H.,
 RA Hosono A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000060; BAA80084.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 103 AA; 10708 MW; 39B6EB5D73DCC267 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 103;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 518 LUGLSTIA 524
 |||||
 DB 53 LUGLSTIA 59

RESULT 29
 ID 005413 PRELIMINARY; PRT; 107 AA.
 AC 005413;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE DNA FOR ORF'S FROM CHROMOSOME XV.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FY1679;
 RX MEDLINE=96021609; PubMed=8533473;
 RA Zumstein E., Pearson B.M., Kalogeropoulos A., Schweizer M.,
 RT "A 29,425 kb segment on the left arm of yeast chromosome XV contains
 RT more than twice as many unknown as known open reading frames.";
 RL Yeast 11:975-986(1995).
 DR EMBL: X83121; CAA58198.1; -
 SQ SEQUENCE 107 AA; 11764 MW; C64C53C4DED241C2 CRC64;

Query Match 0.7%; Score 7; DB 3; Length 107;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 455 LSSULTS 461
 |||||
 DB 42 LSSULTS 48

RESULT 30
 ID 091AB7 PRELIMINARY; PRT; 110 AA.
 AC 091AB7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE RETINAL DEGRADATION SLOW (FRAGMENT).
 OS Brachydonto terio (Zebrafish) (Zebra danio).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 CC Cypriniformes; Cyprinidae; Rasbora; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Suelmann H., Murray B.W., Klein J.,
 RT "Analysis of Ancient Mhc Class III Synteny by Mapping of Orthologous

RT Genes in the Zebrafish.";
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF210644; AAF70449.1; -
 DR InterPro: IPR000830; RDS_ROM.
 DR InterPro: IPR000301; Transmem_4.
 DR Pfam: PF00335; Transmembrane4_1.
 DR PRINTS: PR00218; PERIPHERNRDS.
 FT NON_TER 1 1
 FT NON_TER 110 110
 SQ SEQUENCE 110 AA; 12412 MW; A3799C90373D1580 CRC64;

Query Match 0.7%; Score 7; DB 13; Length 110;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 783 LAELGLKN 789
 |||||
 DB 68 LAELGLKN 74

RESULT 31
 ID 09PGQ1 PRELIMINARY; PRT; 111 AA.
 AC 09PGQ1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF0247.
 GN XF0247.
 OS Xylella fastidiosa.
 CC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OX Xylella.
 NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Ataya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Bioness M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colaudo N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
 RA Garnier J., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Queglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Weidans J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL: AE003878; AAF83060.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 111 AA; 12524 MW; AFC5CB4D03672AD5 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 111;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 RALFORM 15
 |||||
 DB 15 RALFORM 21

RESULT 32

O9IAB8 PRELIMINARY; PRT: 120 AA.

AC O9IAB8; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE RETINYL DEGRADATION SLOW (FRAGMENT).
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Suelmann H., Murray B.W., Klein J.;
 RT "Analysis of Ancient Muc Class III Syntaxin by Mapping of Orthologous
 RT Genes in the Zebrafish."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF210643; AAF70448.1; -
 DR InterPro: IPR000830; RDS_ROM.
 DR InterPro: IPR000301; Transmem. 4.
 DR Pfam: PF00335; transmembrane4; 1.
 DR PRINTS: PR00218; PERIPHERINDS.
 FT NON_TER 1
 FT NON_TER 120
 SQ SEQUENCE 120 AA; 13610 MW; FPD2A67B7F1B2B8D CRC64;

Query Match 0.7%; Score 7; DB 13; Length 120;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 783 LAEGLKN 789
 |||||
 DB 78 LAEGLKN 84

RESULT 33

O9GXU3 PRELIMINARY; PRT: 122 AA.

AC O9GXU3; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOTHETICAL 13.6 KDA PROTEIN (FRAGMENT).
 GN LM12.241.
 OS Leishmania major.
 OC Eukaryota; Eukaryotes; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=FRIDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL390114; CAC02105.1; -
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 122
 SQ SEQUENCE 122 AA; 13604 MW; 034645AA7568C1C6 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 122;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 646 SRAVSLE 652
 |||||
 DB 10 SRAVSLE 16

RESULT 34

O9AUB8 PRELIMINARY; PRT: 122 AA.

AC O9AUB8; 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LECICINE-RICH REPEAT PROTEIN (FRAGMENT).
 GN LRRP.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids; Fabales; Fabaceae; Papilionoideae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. GREENFEAST;
 RA Savenstrand H., Brosche M., Angehagen M., Strid A.;
 RT "Molecular markers for ozone stress isolated by suppression
 RT subtractive hybridisation: specificity of gene expression and
 RT identification of a novel stress-regulated gene."
 RL Plant Cell Environ. 23:689-700(2000).
 DR EMBL; AF137354; AAK19053.1; -
 FT NON_TER 1
 FT NON_TER 122
 SQ SEQUENCE 122 AA; 13805 MW; 00292FF396CFB044 CRC64;

Query Match 0.7%; Score 7; DB 10; Length 122;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 759 LGNLKML 765
 |||||
 DB 55 LGNLKML 61

RESULT 35

O9LB05 PRELIMINARY; PRT: 133 AA.

AC O9LB05; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE ABC TRANSPORTER (FRAGMENT).
 OS Streptomyces noursei.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1971;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=ATCC11455; PubMed=10746764;
 RA MEDLINE=20208553; PubMed=10746764;
 RA Zotchev S., Haugan K., Sekurova O., Sletta H., Ellingsen T.E.,
 RA Valla S.;
 RT "Identification of a gene cluster for antibacterial polyketide-derived
 RT antibiotic biosynthesis in the nystatin producer Streptomyces noursei
 RT ATCC 11455."
 RL Microbiology 146:611-619(2000).
 DR EMBL; AF071512; AAF31484.1; -
 DR InterPro: IPR001687; ATP_GTP_A.
 FT NON_TER 1
 FT NON_TER 133
 SQ SEQUENCE 133 AA; 14443 MW; 2EDCD95595AF0BAE CRC64;

Query Match 0.7%; Score 7; DB 2; Length 133;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180
|||||||
Db 80 GKSTLLQ 86

RESULT 36

O9YB7 PRELIMINARY; PRT; 138 AA.

AC O9YB7; 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)

DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE HYPOHETICAL 14.2 KDA PROTEIN APE1601.

GN APE1601.

OS Aeropyrum pernix.

OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;

OX NCBI_TaxID=56636;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K1;

RX MEDLINE=99310339; PubMed=10382966;

RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,

RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,

RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,

RA Takamaya M., Masuda S., Funahashi T., Tanaka T., Kuboh Y.,

RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,

RT "Complete genome sequence of an aerobic hyper-thermophilic

RT crenarchaeon, Aeropyrum pernix K1."

RT DNA Res. 6:83-101(1999).

DR EMBL: AP000062; BAA0601.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 138 AA; 14230 MW; DEEFEC28D74BB37 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 138;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 686 SATSLRL 692
|||||||
Db 14 SATSLRL 20

RESULT 37

O9G4E5 PRELIMINARY; PRT; 140 AA.

AC O9G4E5; 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).

GN ND1.

OS Fasciola gigantica.

OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;

OC Trematoda; Digenea; Echinostomida; Echinostomata; Fasciolidae;

OC Fasciolidae; Fasciola.

OX NCBI_TaxID=46835;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20436954; PubMed=11040285;

RT van Herwerden L., Blair D., Agatsuma T.;

RT "Multiple lineages of the mitochondrial gene NADH dehydrogenase

RT subunit 1 (ND1) in parasitic helminths: implications for molecular

RT evolutionary studies of facultatively anaerobic eukaryotes."

RT J. Mol. Evol. 51:339-352(2000).

FT NON_TER 1 1
FT NON_TER 140 140
SQ SEQUENCE 140 AA; 15732 MW; 0194D69B3CABBF40 CRC64;

Query Match 0.7%; Score 7; DB 8; Length 140;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KKFVFF 202
|||||||
Db 13 KKFVFF 19

RESULT 38

O9RMH6 PRELIMINARY; PRT; 142 AA.

AC O9RMH6; 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)

DE HMV PROTEIN.

GN HMV.

OS Porphyromonas gingivalis (Bacteroides gingivalis).

OC Bacteria; CF group; Bacteroidaceae; Porphyromonas.

OX NCBI_TaxID=837;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A436;

RA Genco C.A., Wang C.-Y., Simpson W.;

RT "Hmvr, a TonB-dependent receptor required for hemin and hemoglobin

RT utilization in Porphyromonas gingivalis."

RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF200358; AAF07986.1; -

SQ SEQUENCE 142 AA; 15562 MW; 4B774BBA90F40C21 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 142;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 GSGGK 175
|||||||
Db 13 GSGGK 19

RESULT 39

O9HYM6 PRELIMINARY; PRT; 146 AA.

AC O9HYM6; 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE, FKBP-TYPE.

GN PA458.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen."

RT Nature 406:959-964(2000).

DR EMBL: AE004869; AAG07946.1; -

DR InterPro: IPR001179; FKBP_PPIase.
 DR Pfam: PF00254; FKBP_1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; UNKNOWN_1.
 DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
 DR PROSITE: PS00059; FKBP_PPIASE_3; 1.
 KW Isomerase: Complete proteome.
 SQ SEQUENCE 146 AA; 15984 MW; 78BA27DE59C9E599 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 ELAEGCL 322
 |||||
 Db 92 ELAEGCL 98

RESULT 40
 Q9L2N1 PRELIMINARY; PRT: 146 AA.
 AC Q9L2N1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DE HYPOTHEICAL 16.1 KDA PROTEIN.
 GN T7H20_50.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Beyer M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
 RA Bancroft I., Mehes H.W., Rudd S., Lemcke K., Mayer K.F.X.:
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL162508; CAB82973.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 146 AA; 16107 MW; D8C4A1E855D68DE CRC64;

Query Match 0.7%; Score 7; DB 10; Length 146;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 IFSSATS 689
 |||||
 Db 54 IFSSATS 60

RESULT 41
 Q9PHF7 PRELIMINARY; PRT: 157 AA.
 AC Q9PHF7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE HYPOTHEICAL PROTEIN XFA0051.
 GN XFA0051.
 OS Xylella fastidiosa.
 OC Plasmid pXF51.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-9ASC;
 RX MEDLINE-20365717; PubMed-10910347;

RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Coultou N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Factuncani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.U.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidants J., Setubal J.C.,
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."
 RL Nature 406:151-159(2000).
 DR EMBL: AE003851; AAF85619.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 157 AA; 17141 MW; 50698E404D9F5B3A CRC64;

Query Match 0.7%; Score 7; DB 2; Length 157;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 SLSLVIS 709
 |||||
 Db 113 SLSLVIS 119

RESULT 42
 O19028 PRELIMINARY; PRT: 158 AA.
 ID O19028;
 AC O19028;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
 DT 01-AUG-1998 (TREMBLrel. 07, last annotation update)
 DE GLYCOPROTEIN ZONA PELLUCIDA-1 (FRAGMENT).
 GN ZP-1.
 OS Macaca radiata (Bonne monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVARY;
 RA Gupta S.K., Sharma M., Behera A.K., Bisht R., Kaul R.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: Y10382; CAA71410.1; -
 FT NON_TER 158 158
 SQ SEQUENCE 158 AA; 17416 MW; DB14A312E01EC45 CRC64;

Query Match 0.7%; Score 7; DB 6; Length 158;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 SLSLVIS 709
 |||||
 Db 12 SLSLVIS 18

RESULT 43
 09KSD6 PRELIMINARY; PRT; 161 AA.
 ID 09K5D6;
 AC 09K5D6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PUTATIVE GAM PROTEIN.
 OS Campylobacter coli.
 OG plasmid pBT9810.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=195;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 51729;
 RA Dep M.S., Mendz G.L., Smith M.A., Coloe P.J., Fry B.N., Korolik V.;
 RT "Differentiation between Campylobacter hyoilei and Campylobacter coli using genotypic and phenotypic analyses."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ271334; CAB94942.1; -;
 KW Plasmid.
 SQ SEQUENCE 161 AA; 18470 MW; 2662F94D583E78F2 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 161;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 909 LVKGLK 915
 Db 135 LVKGLK 141

RESULT 44
 09KHE3 PRELIMINARY; PRT; 164 AA.
 ID 09KHE3;
 AC 09KHE3;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PUTATIVE ACETYLTRANSFERASE.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC 7120;
 RA Khuyakov I.Y., Golden J.W.;
 RT "Identification of three additional group 2 sigma factor genes in Anabaena sp. strain PCC 7120."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF62218; AAF75760.1; -;
 DR InterPro: IPR000182; Acetyltransf_1.GCNS.
 DR Pfam: PF00583; Acetyltransf_1.
 KW Transferrase.
 SQ SEQUENCE 164 AA; 18580 MW; 954A2C6DEF87DE2C CRC64;

Query Match 0.7%; Score 7; DB 2; Length 164;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 313 LIKELAE 319
 Db 23 LIKELAE 29

RESULT 45
 014843 PRELIMINARY; PRT; 170 AA.
 ID 014843

AC 014843;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MYOSIN LIGHT CHAIN 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKELTAL MUSCLE;
 RA Wu O.L.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -i- SIMILARITY: TO EF-HAND FAMILY.
 DR EMBL; M21812; AAA91848.1; -;
 DR HSSP; P02593; ICDM.
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; ehand; 2.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Calcium-binding; Myosin.
 SQ SEQUENCE 170 AA; 19086 MW; 1E97AC48C411F58F CRC64;

Query Match 0.7%; Score 7; DB 4; Length 170;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1014 VITGAFK 1020
 Db 100 VITGAFK 106

RESULT 46
 028710 PRELIMINARY; PRT; 170 AA.
 ID 028710;
 AC 028710;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MYOSIN LIGHT CHAIN 2.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKELTAL MUSCLE;
 RA Wu O.L.;
 RL Gene 0:0-0(0).
 CC -i- SIMILARITY: TO EF-HAND FAMILY.
 DR EMBL; M21983; AAA91894.1; -;
 DR HSSP; P02593; ICDM.
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; ehand; 2.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Calcium-binding; Myosin.
 SQ SEQUENCE 170 AA; 19040 MW; E70C921B9E900821 CRC64;

Query Match 0.7%; Score 7; DB 6; Length 170;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1014 VITGAFK 1020
 Db 100 VITGAFK 106

RESULT 47
 053048 PRELIMINARY; PRT; 171 AA.
 AC 053048

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF1, PHBCRR, ORP3 AND ORF4.
OS Rhodococcus ruber.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1830;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB40126;
RX MEDLINE=92406022; PubMed=1526467;
RA Pieper U., Steinbuechel A.;
RT "Identification, cloning and sequence analysis of the poly(3-
hydroxyalkanoic acid) synthase gene of the gram-positive bacterium
Rhodococcus ruber.";
RL FEMS Microbiol. Lett. 96:73-80(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB40126;
RX MEDLINE=92406022; PubMed=1526467;
RA Pieper U., Steinbuechel A.;
RT "Identification, cloning and sequence analysis of the poly(3-
hydroxyalkanoic acid) synthase gene of the gram-positive bacterium
Rhodococcus ruber.";
RL FEMS Microbiol. Lett. 75:73-79(1992).
RT Rhodococcus ruber.;
RL FEMS Microbiol. Lett. 75:73-79(1992).
DR InterPro: IPR001387; HTH_3.
DR Pfam: PF01381; HTH_3; 1.
DR SMART: SMO0530; HTH_XRE; 1.
SQ SEQUENCE 171 AA; 18944 MW; ASD7F4B3BF1E08E2 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 292 ALTAEVG 298
Db 112 ALTAEVG 118

RESULT 48
063507 PRELIMINARY; PRT; 176 AA.
AC 063507;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NICOTINAMIDE ADENINE DINUCLEOTIDE DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
GN NDI.
OS Echinostoma sp.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Echinostomida; Echinostomata; Echinostomatidae;
OC Echinostoma.
OX NCBI_TaxID=48218;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP.1. AFRICA;
RA Morgan J.A.T., Blair D.;
RL Parasitology 0:0-0(1997).
DR EMBL: AF025836; AAC16517.1;
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh; 2.
KW Mitochondrion.
FT NON_TER 1
FT NON_TER 176
SQ SEQUENCE 176 AA; 19703 MW; 89FA25FE35692F04 CRC64;

Query Match 0.7%; Score 7; DB 8; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 196 KRFVVF 202
Db 26 KRFVVF 32

RESULT 49
063508 PRELIMINARY; PRT; 176 AA.
AC 063508;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NICOTINAMIDE ADENINE DINUCLEOTIDE DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
GN NDI.
OS Echinostoma caproni.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Echinostomida; Echinostomata; Echinostomatidae;
OX NCBI_TaxID=27848;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADGASCAR (C), EGYP (L);
RA Morgan J.A.T., Blair D.;
RL Parasitology 0:0-0(1997).
DR EMBL: AF025837; AAC16518.1;
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh; 2.
KW Mitochondrion.
FT NON_TER 1
FT NON_TER 176
SQ SEQUENCE 176 AA; 19698 MW; 3826823F9A67F0B2 CRC64;

Query Match 0.7%; Score 7; DB 8; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KRFVVF 202
Db 26 KRFVVF 32

RESULT 50
063509 PRELIMINARY; PRT; 176 AA.
AC 063509;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NICOTINAMIDE ADENINE DINUCLEOTIDE DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
GN NDI.
OS Echinostoma caproni.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Echinostomida; Echinostomata; Echinostomatidae;
OC Echinostoma.
OX NCBI_TaxID=27848;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAMEROON (K);
RA Morgan J.A.T., Blair D.;
RL Parasitology 0:0-0(1997).
DR EMBL: AF025838; AAC16519.1;
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh; 2.
KW Mitochondrion.
FT NON_TER 1
FT NON_TER 176
SQ SEQUENCE 176 AA; 19705 MW; 42ABE4D00C285270 CRC64;

Query Match 0.7%; Score 7; DB 8; Length 176;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KFKFVFF 202
 |||||
 Db 26 KFKFVFF 32

RESULT 51

028975 PRELIMINARY; PRT; 178 AA.
 AC 028975;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AF1294.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OC NCBI_TaxID=2234;
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 Richardson D.L., Kervavage A.R., Graham D.E., Kyriakides N.C.,
 Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 Cotton M.D., Spitzgs T., Artlich P., Kaine B.P., Sykes S.M.,
 Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 Venter J.C.;
 RA "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus."
 RL Nature 390:364-370(1997).
 DR EMBL; AE001015; AAB89954.1; -
 DR TIGR; AF1294; -
 DR InterPro; IPR002839; DUF125.
 DR Pfam; PF01988; DUF125; 1.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 178 AA; 19142 MW; 9910979D341C954E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 178;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 FGALTAE 296
 |||||
 Db 48 FGALTAE 54

RESULT 52

09GXH7 PRELIMINARY; PRT; 179 AA.
 AC 09GXH7;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 18.7 KDA PROTEIN (FRAGMENT).
 GN LM12.392.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 DR STRAIN-FRIEDLIN;

RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL390114; CAC02215.1; -
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 179
 SQ SEQUENCE 179 AA; 18665 MW; D6EE38C85065737A CRC64;

QY 719 VEASPLT 725
 |||||
 Db 21 VEASPLT 27

Query Match

0.7%; Score 7; DB 5; Length 179;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 719 VEASPLT 725
 |||||
 Db 21 VEASPLT 27

RESULT 53

053573 PRELIMINARY; PRT; 181 AA.
 AC 053573;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBLrel. 08, Last annotation update)
 DE ORF C.
 GN PMAB.
 OS Synecchococcus sp.
 OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
 OC NCBI_TaxID=1131;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=1552863;
 RX MEDLINE=92204021;
 RA Der Plas J., Oosterhoff-Theetstra R., Borrias M., Weisbeek P.;
 RT "Identification of replication and stability functions in the complete
 nucleotide sequence of plasmid pUH24 from the cyanobacterium
 RT Synecchococcus sp. PCC 7942."
 RL Mol. Microbiol. 6:653-664(1992).
 DR EMBL; S89470; AAB21872.1; -
 DR SEQUENCE 181 AA; 19535 MW; D2146F13AC5CCD84 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 181;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 878 LEOLTAL 884
 |||||
 Db 66 LEOLTAL 72

RESULT 54

004365 PRELIMINARY; PRT; 188 AA.
 AC 004365;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 21.5 KDA PROTEIN (LIMD-9 REPEITIVE SEQUENCE).
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88062711; PubMed=3681977;
 RA Shehee W.R., Chao S.F., Loeb D.D., Comer M.B., Hutchison C.A. III,
 RA Edgell M.H.;
 RT "determination of a functional ancestral sequence and definition of
 the 5' end of A-type mouse LI elements."
 RL J. Mol. Biol. 196:757-767(1987).
 DR EMBL; M29325; AAA39399.1; -

KW Hypothetical protein.
SQ SEQUENCE 188 AA; 21556 MW; 6CCE7178666793B7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 11; Length 188;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 FLKSLKE 75
|||||||
Db 58 FLKSLKE 64

RESULT 55
Q9KEN8 PRELIMINARY; PRT; 210 AA.
AC Q9KEN8:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ABC TRANSPORTER (ATP-BINDING PROTEIN).
GN BH0814.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
DR EMBL; AP001509; BAB04533.1; -;
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Complete proteome; Transport.
SQ SEQUENCE 210 AA; 22924 MW; 6C73F90C278CEE29 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 GKSTLQ 180
|||||||
Db 10 GKSTLQ 16

RESULT 56
Q9FPJ5 PRELIMINARY; PRT; 218 AA.
AC Q9FPJ5:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE T28A8 30.
OS Arabidopsis thaliana (Mouse-ear cress). Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

RA Chao Q., Brooks S., Chen H., Johnson-Hopson C., Khan S., Shin P.,
RA Ecker J.;
RT "Full length cDNA sequence of Arabidopsis thaliana."
RT Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF324989; AAC40341.1; -;
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR000360; Transketolase.
DR Pfam; PF00560; LRR; 3.
DR PRINTS; PR00019; LEUICHRPT.
DR SMART; SM00370; LRR; 5.
DR PROSITE; PS00801; TRANSKETOLASE_1; UNKNOWN_1.
SQ SEQUENCE 218 AA; 24097 MW; 39B5CB5F85723A39 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 218;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 LGNKNL 765
|||||||
Db 114 LGNKNL 120

RESULT 57
Q9PN21 PRELIMINARY; PRT; 221 AA.
AC Q9PN21:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE PUTATIVE ABC TRANSPORTER ATP-BINDING PROTEIN.
GN Cuj277C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
DR EMBL; AL139077; CAB73531.1; -;
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Complete proteome; Transport.
SQ SEQUENCE 221 AA; 24954 MW; 486539E2B27A947B CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 221;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 LKLROR 242
|||||||
Db 75 LKLROR 81

RESULT 58
O83565


```

ID 083565          PRELIMINARY:      PRT:      222 AA.
AC 083565;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PHOSHOGLYCULATE PHOSPHATASE (GPH-2).
GN TP0554.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
NCBI_TaxID=160;
RX SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9655876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Attlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
RL EMBL: AE001231; AAC26558.1; -.
DR TIGR: TP0554; -.
DR InterPro: IPR001454; Hydrolase.
DR Pfam: PF00702; Hydrolase; 1.
KW Complete proteome.
SQ SEQUENCE 222 AA: 24577 MW: 1EFC03734C5BF980 CRC64;

```

```

Query Match          0.7%; Score 7; DB 2; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 SRALIOR 14
   |||||
Db 52 SRALIOR 58

```

```

RESULT 59
O9VRES          PRELIMINARY:      PRT:      222 AA.
AC O9VRES;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CG15448 PROTEIN.
GN CG15448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RX SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Bottler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshiri A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RL EMBL: AE003569; AAF50856.1; -.
DR FlyBase: FBgn0031153; CG15448.
SQ SEQUENCE 222 AA: 25563 MW: C199AD5D52FE3407 CRC64;

```

```

Query Match          0.7%; Score 7; DB 5; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 703 SLSLVLS 709
   |||||
Db 49 SLSLVLS 55

```

```

RESULT 60
O9YMD9          PRELIMINARY:      PRT:      226 AA.
ID O9YMD9;
AC O9YMD9;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE LMDR-128 Peptide.
OS Lymantria dispar nucleolar polyhedrosis virus (LdMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
NCBI_TaxID=10449;
RX SEQUENCE FROM N.A.
RX MEDLINE=99124785; PubMed=9887315;
RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA Slavicek J.M., Rohmann G.F.;
RT "Sequence and analysis of the genome of a baculovirus pathogenic for
RT Lymantria dispar."
RL Virology 253:17-34(1999).
RN [2]
RX SEQUENCE FROM N.A.
RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA Slavicek J., Rohmann G.F.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RX SEQUENCE FROM N.A.
RA Kuzio J.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF081810; AAC70314.1; -.
SQ SEQUENCE 226 AA: 25461 MW: 772B2FAD174D1DFD CRC64;

```

```

Query Match          0.7%; Score 7; DB 12; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 KSALSOE 574
 |||||
 DB 109 KSALSOE 115

RESULT 61
 09AD75 PRELIMINARY; PRT; 227 AA.

AC 09AD75;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE ABC TRANSPORTER ATP-BINDING PROTEIN.
 GN SCK13.28.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID=1902;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 DE Seeger K.J.; Harris D.;
 RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RL (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 DE Cerdeno A.M.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
 RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RL (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 DE MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M.; Kleser H.M.; Denapate D.; Eichner A.; Cullum J.;
 RA Kinashi H.; Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL512667; CAC21637.2; -.
 KW ATP-binding.
 SQ SEQUENCE 227 AA; 2448 MW; CA88B59FF7FD9FA CRC64;

Query Match 0.7%; Score 7; DB 2; Length 227;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 GKSTLLQ 180
 |||||
 DB 43 GKSTLLQ 49

RESULT 62
 09ULR9 PRELIMINARY; PRT; 229 AA.

AC 09ULR9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE KIA1151 PROTEIN (FRAGMENT).
 GN KIA1151.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 DE MEDLINE=20039618; PubMed=10574461;
 RA Hirosewa M.; Nagase T.; Ishikawa K.; Kikuno R.; Nomura N.; Ohara O.;
 RT "Characterization of cDNA clones selected by the Genekart analysis
 from size-fractionated cDNA libraries from human brain.";

RL DNA Res. 6:329-336(1999).
 DR EMBL; AB032977; BAA86465.1; -.
 FT NON-TER 1
 SQ SEQUENCE 229 AA; 26155 MW; F07C82410E9036B5 CRC64;

Query Match 0.7%; Score 7; DB 4; Length 229;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 MAMLLKL 239
 |||||
 DB 199 MAMLLKL 205

RESULT 63
 095372 PRELIMINARY; PRT; 231 AA.

AC 095372;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ACYL-PROTEIN THIOESTERASE.
 GN DJ886K2.4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Kuznetsov S.R.; Jones T.L.Z.;
 RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RL (2)
 RP SEQUENCE FROM N.A.
 RA Frankland J.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF098668; AAC72844.1; -.
 DR EMBL; AL031295; CAB40158.1; -.
 DR HSP; Q53547; LAUO.
 DR InterPro: IPR000379; Est_11p.thioest_acsite.
 DR InterPro: IPR003140; PIP_Cesterase.
 DR Pfam: PF02230; abhydrolase_2; 1.
 SQ SEQUENCE 231 AA; 24737 MW; 813C9C71757C5135 CRC64;

Query Match 0.7%; Score 7; DB 4; Length 231;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 FGALTAE 296
 |||||
 DB 183 FGALTAE 189

RESULT 64
 09OYL8 PRELIMINARY; PRT; 231 AA.

AC 09OYL8;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LYOPHOSPHOLIPASE II.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Sugimoto H.;
 RT "Rat lysophospholipase II";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB021645; BAA87911.1; -.

DR HSSP: Q53547; 1AUO.
 DR InterPro: IPR000379; Est_lip_thioest_actstle.
 DR InterPro: IPR003140; PLP_Cesterase.
 DR Pfam: PF02230; abhydrolase_2; 1.
 SQ SEQUENCE 231 AA; 24807 MW; 49A710C5A997C7C1 CRC64;

Query Match 0.7%; Score 7; DB 11; Length 231;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 FGALTAE 296
 |||||
 DB 183 FGALTAE 189

RESULT 65
 Q9WTL7 PRELIMINARY; PRT; 231 AA.
 AC Q9WTL7;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LYSOPHOSPHOLIPASE II (LYSOPHOSPHOLIPASE 2).
 GN LYP1A2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheta; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=99165589; PubMed=10664901;
 RT Toyoda T., Sugimoto H., Yamashita S.;
 RT "Sequence expression in *Escherichia coli*, and characterization of
 RT lysophospholipase II.";
 RL Biochim. Biophys. Acta 1437:182-193(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21083660; PubMed=11217851;
 RA Kawai J., Shinaagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Kato T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Botjuna N., Carlini P., de Bonaudo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AB009653; BAA76751.1; -;
 DR EMBL: AK003689; BAB22940.1; -;
 DR HSSP: Q53547; 1AUO.
 DR MGD: MGI:1347000; Lyp1a2.
 DR InterPro: IPR000379; Est_lip_thioest_actstle.
 DR InterPro: IPR003140; PLP_Cesterase.
 DR Pfam: PF02230; abhydrolase_2; 1.
 SQ SEQUENCE 231 AA; 24794 MW; E18797A17570AA97 CRC64;

Query Match 0.7%; Score 7; DB 11; Length 231;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 290 FGALTAE 296
 |||||
 DB 183 FGALTAE 189

RESULT 66
 Q9Z8V6 PRELIMINARY; PRT; 233 AA.
 ID Q9Z8V6;
 AC Q9Z8V6;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE DISULFIDE BOND CHAPERONE
 GN DSBG OR CPN0228 OR CP0536.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CML029;
 RX MEDLINE=9920606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of *Chlamydia pneumoniae* J138
 RT from Japan and CML029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
 RA Berry K., Baas S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Debey R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RT "Genome sequences of *Chlamydia trachomatis* Moyn and *Chlamydia*
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE001608; AAD18381.1; -;
 DR EMBL: AP002545; BAA98438.1; -;
 DR EMBL: AE002213; AAF38359.1; -;
 DR TIGR: CP0536; -;
 KW Complete proteome.
 SQ SEQUENCE 233 AA; 26132 MW; 5EA6542DA8A4ADC4 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 233;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 782 KLAEGLK 788
 |||||
 DB 148 KLAEGLK 154

RESULT 67
 Q9RWN3 PRELIMINARY; PRT; 235 AA.
 ID Q9RWN3;
 AC Q9RWN3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE AZIC FAMILY PROTEIN.

GN DR0633.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1;
RC MEDLINE=2036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffitt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1";
RL Science 286:1571-1577(1999).
DR EMBL; AE001921; AAF10212.1; -.
DR TIGR; DR0633; -.
SQ Complete proteome.
KW SEQUENCE 235 AA; 24153 MW; F6392613F8780785 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 RLPGGLT 756
|111111|
DB 121 RLPGGLT 127

RESULT 68
026543 PRELIMINARY; PRT; 236 AA.
AC 026543.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HYPOTHETICAL 27.4 KDA PROTEIN.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Strigeidida; Schistosomatidae; Schistosomatidae;
OC Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PUERTO RICAN;
RC MEDLINE=93394947; PubMed=7665603;
RA Davis R.E., Hardwick C., Tavernier P., Hodgson S., Singh H.,
RT "RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and
RT genes in the human parasite, Schistosoma mansoni";
RL J. Biol. Chem. 270:21813-21819(1995).
DR EMBL; U30260; AAC46895.1; -.
KW Hypothetical protein.
SQ SEQUENCE 236 AA; 27358 MW; 56ABC24138AA30E6 CRC64;

Query Match 0.7%; Score 7; DB 5; Length 236;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 912 LGLKNMR 918
|111111|
DB 186 LGLKNMR 192

RESULT 69
046699 PRELIMINARY; PRT; 237 AA.
AC 046699;
DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PKM101 CONJUGATION PROTEINS (TRAL), (TRAM), (TRA), (TRAB), (TRAC),
DE (TRAB), (TRAC), (TRAD), (TRAN), (TRAE), (TRAF), (TRAG), ENTRY
DE EXCLUSION PROTEIN (EEX), (KIK), (KOR), (KOR), (KOR), (KOR), (KOR), (KOR),
DE GENES, COMPLETE CDS (TRAM) (TRAB) (TRAB) (TRAD) (TRAE) (TRAF) (EEX)
DE (KOR), (KOR) (NUC).
GN TRAC.
OS Escherichia coli.
OC Plasmid PKM101.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=94302136; PubMed=8029323;
RX Pohlman R.F., Genetti H.D., Winans S.C.;
RA "Entry exclusion of the IncN plasmid PKM101 is mediated by a single
RT hydrophilic protein containing a lipid attachment motif";
RL Plasmid 31:158-165(1994).
RN [2]
RP SEQUENCE FROM N.A.
RP MEDLINE=95198540; PubMed=7891554;
RX Pohlman R.F., Genetti H.D., Winans S.C.;
RA "Common ancestry between IncN conjugal transfer genes and
RT macromolecular export systems of plant and animal pathogens";
RL Mol. Microbiol. 14:655-668(1994).
DR EMBL; U09868; AAA86453.1; -.
KW Plasmid.
SQ SEQUENCE 237 AA; 26055 MW; 94AB09D3E6F3978 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 420 VLLTTGL 426
|111111|
DB 8 VLLTTGL 14

RESULT 70
0916G5 PRELIMINARY; PRT; 237 AA.
AC 0916G5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE TRAC.
OS Escherichia coli.
OC Plasmid PKM101.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RP Eisenbrandt R., Lanka E.;
RL Submitted (FE8-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF236660; AAF40212.1; -.
KW Plasmid.
SQ SEQUENCE 237 AA; 25753 MW; 03DDA05C11BAF0 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 420 VLLTTGL 426
|111111|
DB 8 VLLTTGL 14

RESULT 71

025044 ID 025044 PRELIMINARY: PRT: 240 AA.
 AC 025044:
 DT 01-JUN-1998 (TREMBLrel. 05, Created)
 DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 GN CYTOCHROME C BIOGENESIS PROTEIN (CCDA).
 HP0265.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria: Proteobacteria: epsilon subdivision: Helicobacter group;
 OC Helicobacter.
 OX NCB1_TaxID=210:
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.:
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori."
 RL Nature 388:539-547(1997).
 DR EMBL; AE00545; AAD0731.1; -.
 DR TIGR; HP0265; -.
 DR InterPro: IPR003834; Dsdd_Dipz.
 DR Pfam: PF02683; Dsdd: 1.
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 240 AA; 26522 MW; 13C28C8413FCD422 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 240;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 LFLKSLK 74
 DB 199 LFLKSLK 205
 RESULT 72
 P96389 PRELIMINARY: PRT: 242 AA.
 AC P96389:
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE HYPOHETICAL 26.2 KDA PROTEIN.
 GN RV0207C OR MTCY08D5.02C.
 OS Mycobacterium tuberculosis.
 OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCB1_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulton J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.,
 RT "Deciphering the Biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).

DR EMBL: Z92669; CAB07002.1; -.
 DR TubercuList; RV0207C; -.
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 242 AA; 26175 MW; 315862943884C86E CRC64;

Query Match 0.7%; Score 7; DB 2; Length 242;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 455 LSSLLTS 461
 DB 234 LSSLLTS 240

RESULT 73
 O9CWL7 PRELIMINARY: PRT: 244 AA.
 AC O9CWL7:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 2410018M14RIK PROTEIN.
 GN 2410018M14RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi;
 OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus.
 OX NCB1_TaxID=10090:
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Morone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,
 RA Hayashizaki Y.:
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK010552; BAB27021.1; -.
 DR MGD; MGI:1914234; 2410018M14RIK.
 SQ SEQUENCE 244 AA; 26263 MW; 6F38036D801490AE CRC64;

Query Match 0.7%; Score 7; DB 11; Length 244;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 KKKHKKH 379
 DB 174 KKKHKKH 180

RESULT 74
 O9L0J9 PRELIMINARY: PRT: 246 AA.
 AC O9L0J9:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE PUTATIVE ABC-TRANSPORTER ATP-BINDING PROTEIN.
 GN SCDA0A.12C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cereno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinoshita H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
 CC TRANSPORTERS).
 DR EMBL: AL161691; CAB81857.1; -.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 246 AA; 26164 MW; 2EBFF9753BE0AB62 CRC64;

Query Match 0.7%; Score 7; DB 2; Length 246;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180
 DB 48 GKSTLLQ 54

RESULT 75

ID O9K7R4 PRELIMINARY; PRT; 256 AA.
 AC O9K7R4;
 DT 01-OCT-2000 (TREMBLrel. 15. Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE FERRICHROME ABC TRANSPORTER (ATP-BINDING PROTEIN).
 GN BH3295.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
 CC TRANSPORTERS).
 DR EMBL: AP001518; BAB07014.1; -.

DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Complete proteome; Transport.
 SQ SEQUENCE 256 AA; 29000 MW; 0CD400C506D743AC CRC64;

Query Match 0.7%; Score 7; DB 2; Length 256;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180
 DB 39 GKSTLLQ 45

Search completed: March 25, 2002, 11:05:06
 Job time: 235 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 11:00:51 ; Search time 14.91 Seconds
(without alignments)
2518.094 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 1024
Sequence: 1 MNFKDNRSLIQRMGWTVI.....WQFDDDDLSVTGAFLVTA 1024

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 3664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.9	414	1 SVS2_RAT	P22006 rattus norv
2	9	0.9	508	1 DHAX_PEA	P25795 pisum sativ
3	8	0.8	112	1 RL34_SCHPO	O42846 schizosacch
4	8	0.8	138	1 RYAN_BACSU	P37510 bacillus su
5	8	0.8	225	1 ALKD_PSEPU	P00885 pseudomonas
6	8	0.8	344	1 HMD_METTE	O50758 methanobact
7	8	0.8	344	1 HMD_METTW	P81221 methanobact
8	8	0.8	368	1 SERC_NEIMA	O34370 neisseria m
9	8	0.8	368	1 TYRP_ECOLI	P57007 neisseria m
10	8	0.8	403	1 YI42_MYCTU	O50592 escherichia
11	8	0.8	455	1 ACW2_CHICK	P30372 gallus gall
12	8	0.8	466	1 ACW2_HUMAN	P08172 homo sapien
13	8	0.8	466	1 ACW2_MOUSE	O94244 mus musculu
14	8	0.8	466	1 ACW2_PIG	P06199 sus scrofa
15	8	0.8	466	1 ACW2_RAT	P10980 rattus norv
16	8	0.8	466	1 ACW2_RAT	O64550 rattus norv
17	8	0.8	535	1 YP60_MYCTU	O10771 mycobacteri
18	7	0.7	72	1 YF6J_ECOLI	P76575 escherichia
19	7	0.7	83	1 GRP7_DAUCA	P37704 daucus caro
20	7	0.7	96	1 Y709_RICPR	O92016 rickettsia
21	7	0.7	112	1 POR3_PIG	O29380 sus scrofa
22	7	0.7	118	1 RR19_ASTIO	P34772 astasia lon
23	7	0.7	144	1 GRP1_CHERU	P11898 chenopodium
24	7	0.7	144	1 KBL_SALTY	P37703 daucus caro
25	7	0.7	162	1 MLRS_MOUSE	P04466 rattus norv
26	7	0.7	168	1 MLRS_MOUSE	P04466 rattus norv
27	7	0.7	168	1 MLRS_MOUSE	P04466 rattus norv
28	7	0.7	169	1 MLRT_RABIT	P24732 oryctolagus
29	7	0.7	169	1 MLRT_RABIT	P24732 oryctolagus
30	7	0.7	186	1 Y786_METTA	O58196 methanococc
31	7	0.7	196	1 PH21_PSEFL	O51785 pseudomonas
32	7	0.7	198	1 YH13_YEAST	P38896 saccharomyc
33	7	0.7	198	1 YH13_YEAST	P38896 saccharomyc

34	7	0.7	226	1 OPBD_BACSU	P39775 bacillus su
35	7	0.7	226	1 UCRI_SOLITU	P37841 solanum tub
36	7	0.7	273	1 UCR1_MAIZE	P49727 zea mays (m
37	7	0.7	283	1 POR3_BOVIN	O6m213 bos taurus
38	7	0.7	283	1 POR3_HUMAN	O9y277 homo sapien
39	7	0.7	283	1 POR3_MOUSE	O60931 mus musculu
40	7	0.7	283	1 POR3_RABIT	O9t113 oryctolagus
41	7	0.7	283	1 POR3_RAT	O9r120 rattus norv
42	7	0.7	301	1 Y388_AQUAE	O66709 aquifex aeo
43	7	0.7	305	1 FMT_HELPJ	O92K72 helicobacte
44	7	0.7	308	1 ABCA_AERSA	O07698 aeromonas s
45	7	0.7	308	1 CML1_SCHPO	P53694 schizosacch
46	7	0.7	322	1 ASPG_BACLI	P30363 bacillus 11
47	7	0.7	332	1 OSTR_YEAST	O03723 saccharomyc
48	7	0.7	347	1 ADH_SUISR	P50381 sulfolobus
49	7	0.7	357	1 YCFT_ECOLI	P75955 escherichia
50	7	0.7	359	1 RFI_CHLMU	O9p116 chlamydia m
51	7	0.7	359	1 RFI_CHLTR	O84026 chlamydia t
52	7	0.7	361	1 YB19_SYNY3	P73341 synchocyst
53	7	0.7	378	1 HBCA_SYNY3	P72795 synchocyst
54	7	0.7	383	1 XYLR_STAXY	P27159 staphylococ
55	7	0.7	385	1 Y464_MYCPN	P75112 mycoplasma
56	7	0.7	395	1 SYW_AQUAE	O67115 aquifex aeo
57	7	0.7	396	1 PRRC_ECOLI	P17223 escherichia
58	7	0.7	398	1 KBL_ECOLI	P07912 escherichia
59	7	0.7	402	1 THIB_CANFR	O04677 candida tro
60	7	0.7	407	1 Y097_AQUAE	O66504 aquifex aeo
61	7	0.7	410	1 PGK_PYRMO	P50316 pyrococcus
62	7	0.7	426	1 ZABA_PIG	O29090 s serine/ch
63	7	0.7	431	1 PHOR_ECOLI	P08400 escherichia
64	7	0.7	431	1 PHOR_SHIDY	P45609 shigella dy
65	7	0.7	432	1 GBAL_CRYNE	P54853 cryptococcu
66	7	0.7	439	1 CXA8_SHEEP	P55917 ovis aries
67	7	0.7	447	1 ZABA_HUMAN	O00007 h serine/ch
68	7	0.7	447	1 ZABA_RAT	P36876 r serine/ch
69	7	0.7	453	1 ZABD_RAT	P56932 r serine/ch
70	7	0.7	469	1 LEF1_KLUOLA	P53998 kluyveromyc
71	7	0.7	471	1 Y213_MYCGE	P47455 mycoplasma
72	7	0.7	474	1 LIPL_MOUSE	P11152 mus musculu
73	7	0.7	474	1 LIPL_RAT	O06000 rattus norv
74	7	0.7	482	1 ODB2_BOVIN	P11181 bos taurus
75	7	0.7	482	1 ODB2_HUMAN	P11181 homo sapien
76	7	0.7	482	1 ODB2_MOUSE	P53395 mus musculu
77	7	0.7	494	1 FLAA_HELMO	P50612 helicobacte
78	7	0.7	494	1 PHY5_AVESE	P06595 avena sativ
79	7	0.7	497	1 CBS_DICDI	P46794 dictyosteli
80	7	0.7	501	1 CP7A_RABIT	P51542 oryctolagus
81	7	0.7	503	1 CP7A_RAT	P51825 rattus norv
82	7	0.7	504	1 DEGP_RHIME	O52894 rhizobium m
83	7	0.7	524	1 YB88_YEAST	P38277 saccharomyc
84	7	0.7	529	1 HYAL_CAVPO	P23613 cavia porce
85	7	0.7	529	1 IMA2_HUMAN	P52292 homo sapien
86	7	0.7	529	1 IMA2_MOUSE	P52293 mus musculu
87	7	0.7	531	1 UDI3_RAT	P08430 rattus norv
88	7	0.7	531	1 UDI5_RAT	O64637 rattus norv
89	7	0.7	531	1 UDI5_RAT	O64637 rattus norv
90	7	0.7	531	1 UDI6_HUMAN	P19224 homo sapien
91	7	0.7	531	1 UDI6_HUMAN	O28611 oryctolagus
92	7	0.7	532	1 UDI4_RABIT	O28611 oryctolagus
93	7	0.7	533	1 UDI1_HUMAN	P22309 homo sapien
94	7	0.7	533	1 UDI2_MOUSE	P70691 mus musculu
95	7	0.7	533	1 UDI2_RAT	P20720 rattus norv
96	7	0.7	534	1 UDI3_HUMAN	P35503 homo sapien
97	7	0.7	534	1 UDI4_HUMAN	P22310 homo sapien
98	7	0.7	534	1 UDI5_HUMAN	P35504 homo sapien
99	7	0.7	535	1 UDI1_MOUSE	O63886 mus musculu
100	7	0.7	535	1 YG1S_ECOLI	O46863 escherichia

ALIGNMENTS

RESULT 1

```

SVS2_RAT
ID SVS2_RAT STANDARD: PRT: 414 AA.
AC P22006;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE SEMINAL VESICLE SECRETORY PROTEIN II PRECURSOR (SVS II).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=90277684; PubMed=2351680;
RA Harris S.E., Harris M.A., Johnson C.M., Bean M.F., Dodd J.G.,
RA Matuszk R.J., Carr S.A., Crabb J.W.;
RT "Structural characterization of the rat seminal vesicle secretion II
RT protein and gene."
RL J. Biol. Chem. 265:9896-9903(1990).
CC -1- FUNCTION: THE RAT SEMINAL VESICLE CONTAINS SIX MAJOR ANDROGEN-
CC DEPENDENT SECRETORY PROTEINS REFERRED TO AS SVS I-VI. THE SVS
CC I-III PROTEINS APPEAR TO BE COMPONENTS OF THE RAT COPULATORY
CC PLUG. WITH THE SVS II PROTEIN BEING THE MAJOR COMPONENT.
CC -1- PPM: THE REPEATING UNIT APPEARS TO BE INVOLVED IN THE FORMATION OF
CC THE COPULATORY PLUG VIA A TRANSGLUTAMINASE REACTION CROSS-LINKING
CC GLUTAMINE AND LYSINE RESIDUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05443; AAA42192.1; -.
DR PIR: A36443; A36443.
DR InterPro: IPR002080; SVP-II.
DR PROSITE: PS00515; SVP-II; 12.
KW Testosterone; Seminal vesicle; Signal; Repeat; Copulatory plug.
FT SIGNAL 1 22
FT CHAIN 23 414 SEMINAL VESICLE SECRETORY PROTEIN II.
FT MOD_RES 23 23 PYROLIDONE CARBOXYLIC ACID.
FT DOMAIN 108 311 13 X 13 AA TANDEM REPEATS.
FT REPEAT 108 120 1.
FT REPEAT 127 139 2.
FT REPEAT 140 152 3.
FT REPEAT 153 165 4.
FT REPEAT 166 178 5.
FT REPEAT 179 191 6.
FT REPEAT 192 204 7.
FT REPEAT 205 217 8.
FT REPEAT 224 236 9.
FT REPEAT 237 249 10.
FT REPEAT 257 269 11.
FT REPEAT 275 287 12.
FT REPEAT 299 311 13.
SQ SEQUENCE 414 AA; 45539 MW; C89E3FCE0C0EE92E CRC64;

Query Match 0.9%; Score 9; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1006 QEDDDLSV 1014
DB 335 QEDDDLSV 343

RESULT* 2
DHAX_PEA STANDARD: PRT: 508 AA.
AC P25795;

```

```

DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE TURGOR-RESPONSIVE PROTEIN 26G (EC 1.2.1.-).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euphorbia; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
ON NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. PROGRESS NO. 9;
RX MEDLINE=91355842; PubMed=1715781;
RA Guerrero I., Jones J.T., Mallet J.E.;
RT "Turgor-responsive gene transcription and RNA levels increase rapidly
RT when pea shoots are wilted. Sequence and expression of three
RT inducible genes."
RL Plant Mol. Biol. 15:11-26(1990).
CC -1- INDUCTION: BY DEHYDRATION OF SHOOTS BUT NOT ROOTS AND NOT BY
CC HEAT SHOCK OR ABA.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC SPROGEST, TO MAMMALIAN ANTIOUTRIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X54359; CAA38243.1; -.
DR PIR: S11863; S11863.
DR HSP: P20000; 1A4Z.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh.1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; FALSE_NEG.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 266 266 BY SIMILARITY.
FT ACT_SITE 300 300 BY SIMILARITY.
SQ SEQUENCE 508 AA; 53788 MW; CC88F367B52E923D CRC64;

Query Match 0.9%; Score 9; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 876 NVLEQLTAL 884
DB 315 NVLEQLTAL 323

RESULT* 3
RL34_SCHPO STANDARD: PRT: 112 AA.
AC 042846;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L34.
GN RPL34 OR SPAC23A1.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

```


CC -1- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL021813; CA16982.1; -
 CC InterPro: IPR001284; Ribosomal_L34E.
 CC Pfam: PF01199; Ribosomal_L34E; 1.
 CC PRINTS: PR01250; RIBOSOMAL_L34.
 CC PRODOM: PD005148; RIBOSOMAL_L34E; 1.
 CC PROSITE: PS01145; RIBOSOMAL_L34E; 1.
 CC RIBOSOMAL protein.
 CC SEQUENCE 112 AA; 12870 MW; 9228F904AE223C67 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 112;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 831 CLSANAVK 838
 |||||||
 Db 79 CLSANAVK 86

RESULT 4
 ID YVAN_BACSU STANDARD; PRT; 138 AA.
 AC P37510;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR YVAN.
 OS YVAN.
 ON Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT Systematic sequencing of the 180 kilobase region of the Bacillus
 RT subtilis chromosome containing the replication origin";
 RL DNA Res. 1:1-14(1994).
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D26185; BAA05210.1; -
 CC EMBL: Z99124; CAB16117.1; -
 CC Subtilist: BG10038; yvan.
 CC InterPro: IPR000551; HTH_Merr.
 CC Pfam: PF00376; merr; 1.
 CC SMART: SM00422; HTH_MERR; 1.
 CC PROSITE: PS00552; HTH_MERR_FAMILY; 1.
 CC Hypothetical protein; Transcription regulation; DNA-binding;
 CC Complete proteome.
 FT DNA_BIND 6 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 138 AA; 16048 MW; 63FEB07CE16D1F8B CRC64;

Query Match 0.8%; Score 8; DB 1; Length 138;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 LFLKSLE 75
 |||||||
 Db 51 LFLKSLE 58

RESULT 5
 ID ALKD_PSEPU STANDARD; PRT; 225 AA.
 AC P00885; Q9EV78;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 2-DEHYDRO-3-DEOXYPHOSHOGLUCONATE ALDOLASE (EC 4.1.2.14) (PHOSPHO-2-
 DE DEHYDRO-3-DEOXYGLUCONATE ALDOLASE) (PHOSPHO-2-KETO-3-DEOXYGLUCONATE
 DE ALDOLASE) (KDPG-ALDOLASE).
 GN BDA.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=80159956; PubMed=6988426;
 RA Suzuki N., Wood W.A.;
 RT "Complete primary structure of 2-keto-3-deoxy-6-phosphogluconate
 RT aldolase.";
 RL J. Biol. Chem. 255:3427-3435(1980).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H.
 RA Petruschka L., Burchardt G., Jørgensen J., Adolf K., Herrmann H.;
 RT Analysis of the zwf operon in Pseudomonas putida H.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=83138797; PubMed=7161801;
 RA Mavridis I.M., Hatada M.H., Tulinsky A., Lebloda L.;
 RT "Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8-A
 RT resolution.";
 RL J. Mol. Biol. 162:419-444(1982).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
 RX MEDLINE=77022062; PubMed=974067;
 RA Mavridis I.M., Tulinsky A.;
 RT "The folding and quaternary structure of trimeric 2-keto-3-deoxy-6-
 RT phosphogluconic aldolase at 3.5-A resolution.";
 RL Biochemistry 15:4410-4417(1976).
 CC -1- CATALYTIC ACTIVITY: 2-DEHYDRO-3-DEOXY-D-GLUCONATE 6-PHOSPHATE -
 CC PYRUVATE + D-GLYCERALDEHYDE 3-PHOSPHATE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: BELONGS TO THE KHG/KDPG ALDOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: A1279003; CAC14910.1; -
 CC PIR: A01105; ADPSGP.
 CC PDB: 1KGA; 17-FEB-84.
 CC InterPro: IPR000887; Aldolase_KDPG.
 CC Pfam: PF01081; Aldolase; 1.
 CC PROSITE: PS00159; ALDOLASE_KDPG_KHG; 1.
 CC PROSITE: PS00160; ALDOLASE_KDPG_KHG_2; 1.
 CC Lyase; Schiff base; 3D-structure.

```

DE INT_MET 0
DE ACT_SITE 60
DE BINDING 144
DE CONFLICT 109
DE CONFLICT 194
DE CONFLICT 219
DE SEQUENCE 225 AA; 23937 MW; 2573E976734A4757 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

660 RTLEVTLR 667
|||||
53 RTLEVTLR 60

RESULT 6
HMD_METTW
ID HMD_METTW STANDARD; PRT; 344 AA.
AC 050758;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE H(2)-FORMING N5,N10-METHYLENETETRAHYDROMETHANOPTERIN DEHYDROGENASE
DE (EC 1.5.99.11) (H2-DEPENDENT METHYLENE-H4MPT DEHYDROGENASE).
GN HMD.
OS Methanobacterium thermoformicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3664 / CB12;
RA Vaupel M.;
RT Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: N(5),N(10)-METHYLENETETRAHYDROMETHANOPTERIN +
CC REDUCED COENZYME F(420) = 5-METHYL-5,6,7,8-TETRAHYDROMETHANOPTERIN
CC + COENZYME F(420).
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- PATHWAY: INVOLVED IN METHANOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; X92711; CAA63376.1; -.
KM Oxidoreductase; Methanogenesis; Zinc.
SQ SEQUENCE 344 AA; 37525 MW; 33C6E5C0D5C030B7 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

878 LEOLTALM 885
|||||
292 LEOLTALM 299

RESULT 7
HMD_METTW
ID HMD_METTW STANDARD; PRT; 344 AA.
AC P81221; O50526;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE H(2)-FORMING N5,N10-METHYLENETETRAHYDROMETHANOPTERIN DEHYDROGENASE

```

```

DE (EC 1.5.99.11) (H2-DEPENDENT METHYLENE-H4MPT DEHYDROGENASE).
GN HMD OR MTH.
OS Methanobacterium thermoautotrophicum (strain Winter).
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=79930;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95247681; PubMed=7730278;
RA Nollig J., Phil T.D., Vriesema A., Reeve J.N.;
RT "Organization and growth phase-dependent transcription of methane
RT genes in two regions of the Methanobacterium thermoautotrophicum
RT genome.";
RL J. Bacteriol. 177:2460-2468(1995).
CC -1- CATALYTIC ACTIVITY: N(5),N(10)-METHYLENETETRAHYDROMETHANOPTERIN +
CC REDUCED COENZYME F(420) = 5-METHYL-5,6,7,8-TETRAHYDROMETHANOPTERIN
CC + COENZYME F(420).
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- PATHWAY: INVOLVED IN METHANOGENESIS.
CC -1- SUBUNIT: HOMODIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; U19364; AA87437.1; -.
KM Oxidoreductase; Methanogenesis; Zinc.
SQ SEQUENCE 344 AA; 37539 MW; 33C6E5C75C390B1 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

878 LEOLTALM 885
|||||
292 LEOLTALM 299

RESULT 8
SERC_NEIMA
ID SERC_NEIMA STANDARD; PRT; 368 AA.
AC O34370; O33382; O33383; O33384; O33386;
DT 15-DEC-1998 (Rel. 37, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52) (PSAT).
GN SERC OR NMA1894.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Mueller K., Selter A., Wang J.-F.,
RA del Valle J., Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during
RT 30 years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jørgensen K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

```

```

RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
  meningitidis 22491."
RL Nature 404:502-506(2000).
CC -1- CATALYTIC ACTIVITY: O-PHOSPHO-L-SERINE + 2 OXOGLUTARATE =
CC 3-PHOSPHONOOXYPYRUVATE + L-GLUTAMATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE
CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL: AF004820: AAC32675.1; -
CC EMBL: AF004821: AAC32679.1; -
CC EMBL: AF004822: AAC32683.1; -
CC EMBL: AF004823: AAC32687.1; -
CC EMBL: AF004824: AAC32691.1; -
CC EMBL: AF004825: AAC32695.1; -
CC EMBL: AF004826: AAC32699.1; -
CC EMBL: AF004827: CAB85115.1; -
CC InterPro: IPR000192; AminoTransf_class_V.
CC InterPro: IPR003248; Phosphser_aminotransf.
CC Pfam: PF00266; aminotran_5; 1.
CC ProDom: PD001544; Phosphser_aminotransf. 1.
CC PROSITE: PS00595; AA_TRANSFERS_CLASS_5; 1.
CC Serine biosynthesis; Pyridoxine biosynthesis; Transferase;
CC AminoTransferase; Pyridoxal phosphate; Complete proteome.
CC FT BINDING 203 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC FT VARIANT 168 168 R -> C (IN STRAINS B293, Z3910 AND
CC Z3918).
CC FT VARIANT 192 192 A -> S (IN STRAIN Z3524).
CC FT VARIANT 237 237 I -> L (IN STRAINS B293, Z3524, Z3910,
CC Z3915 AND Z3918).
CC FT VARIANT 240 240 D -> E (IN STRAINS Z3915 AND Z3524).
CC FT VARIANT 289 289 G -> D (IN STRAINS B293, Z3910 AND
CC Z3918).
CC FT VARIANT 336 336 T -> S (IN STRAIN Z4296).
CC FT SEQUENCE 368 AA: 41388 MW: 3DBE305853698537 CRC64;
SQ

```

Query Match 0.8%; Score 8; DB 1; Length 368;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 RUTDTEIR 925
 Db 114 RUTDTEIR 121

```

RESULT 9
SERC_NEIMB STANDARD; PRT; 368 AA.
AC P57007;
DT 20-AUG-2001 (Rel. 40; Created)
DT 20-AUG-2001 (Rel. 40; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52) (PSAT).
GN SERC OR NMB1640.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;

```

```

RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson M.C., Gwyn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
  MC58."
RL Science 287:1809-1815(2000).
CC -1- CATALYTIC ACTIVITY: O-PHOSPHO-L-SERINE + 2 OXOGLUTARATE =
CC 3-PHOSPHONOOXYPYRUVATE + L-GLUTAMATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE.
CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL: AE002514; AAF41989.1; -
CC TIGR: NMB1640; -
CC InterPro: IPR000192; AminoTransf_class_V.
CC InterPro: IPR003248; Phosphser_aminotransf.
CC Pfam: PF00266; aminotran_5; 1.
CC ProDom: PD001544; Phosphser_aminotransf. 1.
CC PROSITE: PS00595; AA_TRANSFERS_CLASS_5; 1.
CC Serine biosynthesis; Pyridoxine biosynthesis; Transferase;
CC AminoTransferase; Pyridoxal phosphate; Complete proteome.
CC FT BINDING 203 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC FT SEQUENCE 368 AA: 41393 MW: 97DFCE52BBE5E021 CRC64;
SQ

```

Query Match 0.8%; Score 8; DB 1; Length 368;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 RUTDTEIR 925
 Db 114 RUTDTEIR 121

```

RESULT 10
TYRP_ECOLI STANDARD; PRT; 403 AA.
AC P18199; P76309;
DT 01-NOV-1990 (Rel. 16; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE TYROSINE-SPECIFIC TRANSPORT PROTEIN (TYROSINE PERMEASE).
GN TYRP OR B1907.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008121; PubMed=3049553;
RT "DNA sequence of the gene (tyrP) encoding the tyrosine-specific
  transport system of Escherichia coli."
RL J. Bacteriol. 170:4946-4949(1988).
RN [2]
RP REVISIONS.

```

RX MEDLINE-91216998; PubMed-2022620;
 RA Howlitt S.;
 RL Unpublished results, cited by:
 RL Sarsero J.P., Wookley P.J., Gollnick P., Yanofsky C., Pittard A.J.;
 RL J. Bacteriol. 173:3231-3234(1991).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-97251358; PubMed-9097040;
 RA Itoh T., Alta H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;
 RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 4.0.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 CC -1- FUNCTION: INVOLVED IN TRANSPORTING TYROSINE ACROSS THE CYTOPLASMIC
 CC MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- INDUCTION: REPRESSION BY TYROSINE AND INDUCED BY PHENYLALANINE
 CC -1- SIMILARITY: BELONGS TO THE MTR / TNAH / TYRO PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M23240. AAA24705.1; -;
 DR EMBL; AE000284; AAC74977.1; -;
 DR EMBL; D90832; BAA15730.1; -;
 DR PIR; JS0146; GRECY.
 DR Ecogene; EG11041; TYRP.
 DR InterPro; IPR000293; AA_rel_permease.1.
 DR InterPro; IPR002422; AA_rel_permease.2.
 DR InterPro; IPR002091; Aromatic_AA_permease.
 DR InterPro; IPR003040; Tyr_transport_permease.
 DR PRINTS; PR00166; AROMAPRMEASE.
 DR PRODOM; PD128593; Tyr_transport_permease.1.
 DR PROSITE; PS00594; AROMATIC_AA_PERMEASE.1;
 DR Transport; Amino-acid transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 9 29 POTENTIAL.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 275 295 POTENTIAL.
 FT TRANSMEM 308 328 POTENTIAL.
 FT TRANSMEM 332 352 POTENTIAL.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 131 135 AGGV -> RRYAWL (IN REF. 1).
 FT CONFLICT 222 224 SAI -> ECD (IN REF. 1).
 FT SEQUENCE 403 AA; 42819 MW; 7BCCAA833679796 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 403;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 154 LINGLQAL 161
 |||||
 Db 255 LINGLQAL 262
 RESULT 11
 Y142_MYCTU STANDARD; PRT; 455 AA.
 ID 050592; P95166;
 AC 01-NOV-1997 (Rel. 35, Last Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 48.1 KDA PROTEIN RV1842C.
 GN RV1842C OR MT1890 OR MYCYL11.02 OR MYCY359.31.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA MEDLINE-96259687; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Felschman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Bishai W.;
 RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE UPF0053 FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z83859; CAB06112.1; -;
 DR EMBL; AE007047; AAK46161.1; -;
 CC CC
 DR TIGR; MT1890; -;
 DR TubercuList; RV1842C; -;
 DR InterPro; IPR000644; CBS.
 DR InterPro; IPR002550; DUF21.
 DR Pfam; PF00571; CBS; 2.
 DR Pfam; PF01595; DUF21.1.
 DR SMART; SM00116; CBS; 1.
 DR Hypothetical protein; CBS domain; Repeat; Transmembrane;
 KW Complete proteome.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.

FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT CONFLICT 389 389 G -> S (IN REF. 2).
 SO SEQUENCE 455 AA; 48124 MW; 49F6A31F75166CC CRC64;

Query Match 0.8%; Score 8; DB 1; Length 455;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 289 OFGALTAE 296
 |||||||
 Db 214 OFGALTAE 221

RESULT 12
 ACM2_CHICK STANDARD; PRT; 466 AA.

AC P30372;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MUSCARINIC ACETYLCHOLINE RECEPTOR M2.
 GN CHRM2 OR CM2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A.

RA Tietje K.M., Nathanson N.M.:
 RT "Embryonic chick heart expresses multiple muscarinic acetylcholine
 receptor subtypes. Isolation and characterization of a gene encoding
 a novel m2 muscarinic acetylcholine receptor with high affinity for
 pirenzepine."
 RT J. Biol. Chem. 266:17382-17387(1991).

CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHODIESTERASES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLYLATE CYCLASE INHIBITION.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- MISCELLANEOUS: THIS RECEPTOR HAS A HIGH AFFINITY FOR PIRENZEPINE.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL: M73217; AAB04106.1; -
 DR PIR: A40972; A40972.

DR GCRDB: GCR_0165;
 DR InterPro: IPR000276; GPCR_Rhodopsn.

DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PR00237; GPCRHHODOPSN.

DR PRINTS: PR00243; MUSCARINIC.
 DR PRINTS: PR00539; MUSCARINICM2.

DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_FL_2; 1.

KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor.

FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 26 48 1 (POTENTIAL).
 FT DOMAIN 49 62 1 (POTENTIAL).
 FT TRANSMEM 63 83 2 (POTENTIAL).
 FT DOMAIN 84 100 2 (POTENTIAL).
 FT TRANSMEM 101 122 3 (POTENTIAL).

FT DOMAIN 123 142 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 143 165 4 (POTENTIAL).
 FT DOMAIN 166 187 4 (POTENTIAL).
 FT TRANSMEM 188 210 5 (POTENTIAL).
 FT DOMAIN 211 388 5 (POTENTIAL).
 FT TRANSMEM 389 409 6 (POTENTIAL).
 FT DOMAIN 410 423 7 (POTENTIAL).
 FT TRANSMEM 424 443 7 (POTENTIAL).
 FT DOMAIN 444 466 7 (POTENTIAL).
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 99 179 BY SIMILARITY.
 FT MOD_RES 446 446 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 450 450 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 465 465 PHOSPHORYLATION (POTENTIAL).
 SO SEQUENCE 466 AA; 51565 MW; 2E3D8241D6168995 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 466;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSLV 707
 |||||||
 Db 32 VAGSLSLV 39

RESULT 13
 ACM2_HUMAN STANDARD; PRT; 466 AA.

AC P08172;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MUSCARINIC ACETYLCHOLINE RECEPTOR M2.
 GN CHRM2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.

RA MEDLINE-87263421; PubMed-3037705;
 RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.:
 RT "Identification of a family of muscarinic acetylcholine receptor
 genes."
 RT Science 237:527-532(1987).

CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHODIESTERASES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLYLATE CYCLASE INHIBITION.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).


```

RA Numa S., Haga K., Ichiyama A., Kangawa K., Matsuo H., Hirose T.,
RA Hama T.
RT "Primary structure of porcine cardiac muscarinic acetylcholine
RT receptor deduced from the cDNA sequence."
RL FEBS Lett. 209:367-372(1986).
RM [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87206169; PubMed=3107123;
RA Peralta E.G., Winslow J.W., Peterson G.L., Smith D.H.,
RA Ashkenazi A., Ramachandran J., Schimerlik M.I., Capon D.J.;
RT "Primary structure and biochemical properties of an M2 muscarinic
RT receptor."
RL Science 236:600-606(1987).
CC -I- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
CC BREAKDOWN OF PHOSPHOLIPIDS & MODULATION OF POTASSIUM CHANNELS
CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
CC ADENYLATE CYCLASE INHIBITION.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04708; AAA28413.1; -.
DR EMBL: M16311; AAA30986.1; -.
DR PIR: A27386; A27386.
DR PIR: A25656; A25656.
DR GCRDb: GCR_0102; -.
DR GCRDb: GCR_0105; -.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1
DR PRINTS: PR00237; GPCRHHODPSN.
DR PRINTS: PR00243; MUSCARINICR.
DR PRINTS: PR00539; MUSCRINICM2R.
DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_FL_2; 1.
KM Postsynaptic membrane; Ionic channel family; G-protein coupled receptor.
KW Phosphorylation; Multigene family;
FT DOMAIN 1 22
FT TRANSMEM 23 45
FT DOMAIN 46 59
FT TRANSMEM 60 80
FT DOMAIN 81 97
FT TRANSMEM 98 119
FT DOMAIN 120 139
FT TRANSMEM 140 162
FT DOMAIN 163 184
FT TRANSMEM 185 207
FT DOMAIN 208 288
FT TRANSMEM 389 409
FT DOMAIN 410 423
FT TRANSMEM 424 443
FT DOMAIN 444 466
FT CARBOHYD 2 2
FT CARBOHYD 3 3
FT CARBOHYD 6 6
FT DISULFD 6 176
FT MOD_RES 446 446
FT MOD_RES 450 450
FT MOD_RES 465 465
FT CONFLICT 330 330
FT SEQUENCE 466 AA; 51673 MW; 53D089F19789CD9 CRC64;
Query Match 0.8%; Score 8; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	700	VAGSLSTLV	707	
Db	29	VAGSLSTLV	36	
		RESULT_16		
ID	ACM2_RAT	STANDARD:	PRT:	466 AA.
AC	P10980; Q9Z221;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MUSCARINIC ACETYLCHOLINE RECEPTOR M2.			
CN	CHRM2 OR CHRM-2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
FX	MEDLINE=68068581; PubMed=2825184;			
RA	Goodyne J.D., Robinson D.A., Fitzgerald M.G., Chung F.-Z.,			
RA	Kerlavage A.R., Lentes K.-U., Lai J., Wang C.-D., Fraser C.M.,			
RA	Venter J.C.;			
RT	"Primary structure of rat cardiac beta-adrenergic and muscarinic			
RT	cholinergic receptors obtained by automated DNA sequence analysis:			
RT	further evidence for a multigene family."			
RL	Proc. Natl. Acad. Sci. U.S.A. 84:8296-8300(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Iris;			
RA	MEDLINE=99138467; PubMed=9972520;			
RA	Furuta M., Ohya S., Imaizumi Y., Watanabe M.;			
RT	"Molecular cloning of m3 muscarinic acetylcholine receptor in rat			
RT	Iris."			
RL	J. Smooth Muscle Res. 34:111-122(1998).			
CC	-1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS			
CC	CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,			
CC	BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS			
CC	THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS			
CC	ADENYLATE CYCLASE INHIBITION.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J03025; AAA40926.1; -			
DR	EMBL; AB017655; BAA36838.1; -			
DR	PIR; S10856; S10856.			
DR	GCRdb; GCR_0129; -			
DR	InterPro: IPR000276; GPCR_Rhodpsn.			
DR	Pfam: PF00001; Tfm1.1; 1.			
DR	PRINTS; PR00237; GPCRHOOPS.			
DR	PRINTS; PR00243; MUSCARINICR.			
DR	PRINTS; PR00339; MUSCRINICM2R.			
DR	PROSITE; PS00237; G-PROTEIN_RECP_F1.1; 1.			
DR	PROSITE; PS0262; G-PROTEIN_RECP_F1.2; 1.			
KW	Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;			
KW	phosphorylation; Multigene family; G-protein coupled receptor.			
FT	DOMAIN 1 22			
FT	TRANSMEM 23 45			
FT	DOMAIN 46 59			
FT	TRANSMEM 60 80			
FT	DOMAIN 81 97			
FT	TRANSMEM 98 119			
FT	DOMAIN 120 139			
FT	CYTOPLASMIC (POTENTIAL).			


```

FT TRANSMEM 140 162 4 (POTENTIAL).
FT DOMAIN 163 187 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 185 207 5 (POTENTIAL).
FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 409 6 (POTENTIAL).
FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 424 443 7 (POTENTIAL).
FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 6 6 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 96 176 BY SIMILARITY.
FT MOD_RES 446 446 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 450 450 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 465 465 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 51 51 N -> S (IN REF. 1).
FT CONFLICT 273 273 N -> T (IN REF. 1).
FT CONFLICT 289 290 VS -> SA (IN REF. 1).
FT CONFLICT 313 313 G -> D (IN REF. 1).
FT CONFLICT 337 337 C -> Y (IN REF. 1).
FT CONFLICT 353 353 N -> S (IN REF. 1).
FT CONFLICT 360 360 I -> V (IN REF. 1).
FT CONFLICT 369 369 T -> P (IN REF. 1).
SQ SEQUENCE 466 AA: 51539 MW: 706CCD8636A676B CRC64:

```

```

Query Match 0.8%; Score 8; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 760 VAGSLSLV 707
Db 29 VAGSLSLV 36

```

```

RESULT 17
UD11_RAT STANDARD: PRT: 535 AA.
AC 064550; 064635;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UDP-GLUCURONOSYLTRANSFERASE 1-1 PRECURSOR, MICROSOMAL (EC 2.4.1.17)
DE (UDPgt) (UGT1*1) (UGT1-01) (UGT1.1) (UGT1.1) (B1).
GN UGT1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MISTAR;
RC MEDLINE=95327065; PubMed=7603447;
RA Coffman B. L., Green M.D., King C.O., Teply T.R.;
RT "Cloning and stable expression of a cDNA encoding a rat liver UDP-
RT glucuronosyltransferase (UDP-glucuronosyltransferase 1.1) that
RT catalyzes the glucuronidation of opioids and bilirubin."
RL Mol. Pharmacol. 47:1101-1105(1995).
RN [2]
RN SEQUENCE OF 1-290 FROM N.A.
RP STRAIN=MISTAR;
RC MEDLINE=95322265; PubMed=7608130;
RA Emi Y., Ikushiro S.I., Iyanagi T.;
RT "Drug-responsive and tissue-specific alternative expression of
RT multiple first exons in rat UDP-glucuronosyltransferase family 1
RT (UGT1) gene complex."
RL J. Biochem. 117:392-399(1995).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENOGENOUS COMPOUNDS. GLUCURONATES OPIOIDS AND BILIRUBIN.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR = UDP + ACCEPTOR
CC BETA-D-GLUCURONOSIDE.
CC -1- SUBCELLULAR LOCATION: MICROSOMAL.

```

```

CC -1- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES FOR MANY DIFFERENT
CC UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMINAL DOMAIN AND A
CC COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U20551; AAC52219.1; -
DR EMBL: D38065; BAA07260.1; -
DR InterPro: IPR002213; UDPGT.
DR Pfam: PF00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome; Alternative splicing.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 535 UDP-GLUCURONOSYLTRANSFERASE 1-1.
FT TRANSMEM 493 509 POTENTIAL.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 253 253 A -> D (IN REF. 2).
SQ SEQUENCE 535 AA: 59662 MW: 26B642FCA7DD4082 CRC64:

```

```

Query Match 0.8%; Score 8; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 570 ALSQFEFA 577
Db 287 ALSQFEFA 294

```

```

RESULT 18
YF60_MYCTU STANDARD: PRT: 72 AA.
ID YF60_MYCTU
AC 010771;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 8.2 KDA PROTEIN RV1560.
GN RV1560 OR MT1611 OR MTCY48.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RC MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

```


RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.:
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains",
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO M.TUBERCULOSIS RV2009.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z74020; CAA9833.1; -
 DR EMBL: AE007027; AAK45878.1; -
 DR TIGR: MT1611; -
 DR TubercuList; RV1560; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 72 AA; 8241 MW; FCCFA7E0223701B9 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 72;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 EGDLDL 101
 |||||
 Db 60 EGDLDL 66

RESULT 19
 YFGJ_ECOLI ID STANDARD; PRT; 83 AA.
 AC P76575;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOHETICAL 9.2 KDA PROTEIN IN XSEB-HISS INTERGENIC REGION.
 GN YFGJ OR B2510.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; Pubmed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.:
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000337; AAC75563.1; -
 DR Ecogene; EG14206; Yfgu.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 83 AA; 9193 MW; 60A6A5843F130FDD CRC64;

Query Match 0.7%; Score 7; DB 1; Length 83;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 631 TGGTME 637
 |||||
 Db 8 TGGTME 14

RESULT 20
 GRP7_DAUCA ID STANDARD; PRT; 96 AA.
 AC P37704;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE GLYCINE-RICH PROTEIN DC7.1 PRECURSOR.
 OS Daucus carota (Carrot).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
 OX NCBI_Taxid=4039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aleich F., Richter G.:
 RT "Gene expression during induction of somatic embryogenesis in carrot
 RT cell suspensions.";
 RL Planta 183:17-24(1990).
 CC -1- FUNCTION: MAY BE CONNECTED WITH THE INITIATION OF EMBRYOGENESIS
 CC OR WITH THE METABOLIC CHANGES PRODUCED BY THE REMOVAL OF AUXINS.
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED DURING EARLY
 CC EMBRYOGENESIS.
 CC -1- INDUCTION: BY THE REMOVAL OF AUXINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X15706; CAA33736.1; -
 DR PTR; S35715; S35715.
 KW Repeat; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 96 GLYCINE-RICH PROTEIN DC7.1.
 FT DOMAIN 42 67 2 APPROXIMATE REPEATS OF H-H-G(4,6)-H.
 FT REPEAT 42 50 1.
 FT REPEAT 61 67 2.
 FT REPEAT 61 67 2.
 SQ SEQUENCE 96 AA; 9319 MW; 7C00D44637B7A364 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 96;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 518 LGLSIA 524
 |||||
 Db 8 LGLSIA 14

RESULT 21
 Y709_RICPR ID STANDARD; PRT; 98 AA.
 AC Q9ZCL6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOHETICAL PROTEIN RP709.
 GN RP709.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=782;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sickeritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ235273; CAA1514.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 98 AA; 10996 MW; 24C7EAA1327CF517 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 98;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 625 EKAEDT 631
 |||||
 Db 36 EKAEDT 42

RESULT 22

POR3_PIG STANDARD; PRT; 112 AA.
 AC Q29380; 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (OUTER
 DE MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3) (FRAGMENT).
 GN VDAC3.
 OS Sus scrofa (Pig)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=96823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Small intestine;
 RX MEDLINE=96327607; PubMed=8672129;
 RA Wintecoe A.K., Fredholm M., Davies W.;
 RT "Evaluation and characterization of a porcine small intestine cDNA
 RT library: analysis of 839 clones."
 RL Mamm. Genome 7:509-517(1996).
 CC -----
 CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
 CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING STIDED BETA-SHEETS.
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: F14590; CAA23141.1; -
 DR InterPro: IPR001925; Euk_porin.
 DR Pfam: PF01459; Euk_porin; 1.
 DR PROSITE: PS00558; EUKARYOTIC_PORIN; PARTIAL.
 KW Outer membrane; Porin; Mitochondrion.
 FT NON_TER 1
 FT NON_TER 112

SQ SEQUENCE 112 AA; 12126 MW; 230825D4811593F CRC64;

Query Match 0.7%; Score 7; DB 1; Length 112;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAGLK 788
 |||||
 Db 86 KLAGLK 92

RESULT 23

RI9_ASTLO STANDARD; PRT; 118 AA.
 ID RI9_ASTLO
 AC P34772;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S19.
 DE RPS19.
 GN Astasia longa (Euglenophyceae alga).
 OS Chloroplast.
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Astasia.
 OX NCBI_TaxID=3037;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CCAP 1204-17A;
 RX MEDLINE=95062725; PubMed=7972503;
 RA Gockel G., Baier S., Hachetel W.;
 RT "Plastid ribosomal protein genes from the nonphotosynthetic
 RT flagellate Astasia longa."
 RL Plant Physiol. 105:1443-1444(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CCAP 1204-17A;
 RX MEDLINE=21080522; PubMed=11212895;
 RA Gockel G., Hachetel W.;
 RT "Complete gene map of the plastid genome of the nonphotosynthetic
 RT euglenoid flagellate Astasia longa."
 RL Probst 151:347-351(2000).
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ294725; CAC24597.1; -
 DR PIR: S38608; S38608.
 DR Mendel: 4160; ASTLO; rps19.1.
 DR InterPro: IPR002222; Ribosomal_S19.
 DR Pfam: PF00203; Ribosomal_S19; 1.
 DR PRINTS: PR00975; RIBOSOMALS19.
 DR ProDom: PD001012; Ribosomal_S19; 1.
 DR PROSITE: PS00323; RIBOSOMAL_S19; 1.
 KW Ribosomal protein; Chloroplast.
 SQ SEQUENCE 118 AA; 13664 MW; B84D6141BB2C0E0E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 118;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 963 NKLQVF 969
 |||||
 Db 26 NKLQVF 32

RESULT 24

```

GRP1_CHERU
ID GRP1_CHERU STANDARD: PRT: 144 AA.
AC p11898:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE GLYCINE-RICH PROTEIN HCL.
OS Chenopodium rubrum (Red goosefoot) (Pigweed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllales; Chenopodiaceae; Chenopodium.
OX NCBI_TaxID=3560;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89240041; PubMed=2717413;
RA Kaldenhof R., Richter G.;
RT "Sequence of cDNA for a novel light-induced glycine-rich protein.";
RL Nucleic Acids Res. 17:2853-2853(1989).
CC -1- INDUCTION: BY LIGHT.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: X14067; CAA32230.1; -.
DR PIR: S04069; S04069.
KW Repeat; Transmembrane.
FT TRANSMEM 5 25
FT DOMAIN 37 113
FT REPEAT 37 42
FT REPEAT 43 48
FT REPEAT 50 55
FT REPEAT 56 61
FT REPEAT 63 68
FT REPEAT 69 74
FT REPEAT 76 81
FT REPEAT 82 87
FT REPEAT 89 94
FT REPEAT 102 107
FT REPEAT 108 113
FT REPEAT 144 AA; 14137 MW; 5B4D62D4A61621B0 CRC64;
SQ SEQUENCE 144 AA; 14137 MW; 5B4D62D4A61621B0 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 LIGLISA 524
DB 8 LIGLISA 14

RESULT 25
GRP9_DAUCA STANDARD: PRT: 144 AA.
AC p37703;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE GLYCINE-RICH PROTEIN DC9.1.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
[1]
RN RN
RP SEQUENCE FROM N.A.
RA Aleith F., Richter G.;

```

```

RT "Gene expression during induction of somatic embryogenesis in carrot
RT cell suspensions.";
RL Planta 183:17-24(1990).
DR PIR: S35716; S35716.
KW Repeat; Transmembrane.
FT TRANSMEM 5 25
FT DOMAIN 37 113
FT REPEAT 37 42
FT REPEAT 43 48
FT REPEAT 50 55
FT REPEAT 56 61
FT REPEAT 63 68
FT REPEAT 69 74
FT REPEAT 76 81
FT REPEAT 82 87
FT REPEAT 89 94
FT REPEAT 102 107
FT REPEAT 108 113
FT REPEAT 144 AA; 14111 MW; 5B4D62CFB8A791B0 CRC64;
SQ SEQUENCE 144 AA; 14111 MW; 5B4D62CFB8A791B0 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 LIGLISA 524
DB 8 LIGLISA 14

RESULT 26
KBL_SALTY STANDARD: PRT: 162 AA.
ID KBL_SALTY
AC p37419;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 2-AMINO-3-KETOGLUTARATE COENZYME A LIGASE (EC 2.3.1.29) (AKB LIGASE)
DE (GLYCINE ACETYLTRANSFERASE) (FRAGMENT).
GN KBL.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN RN
RP SEQUENCE FROM N.A.
RX STRAIN=LT2;
RX MEDLINE=94209240; PubMed=8157607;
RA Srisena D.M., MacLachlan P.R., Liu S.L., Hessel A.,
RA Sanderson K.E.;
RT "Molecular analysis of the rfaD gene, for heptose synthesis, and the
RT rfaF gene, for heptose transfer, in lipopolysaccharide synthesis in
RT Salmonella typhimurium.";
RL J. Bacteriol. 176:2379-2385(1994).
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + GLYCINE = COA + 2-AMINO-
CC 3-OXOBUTANOATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: U06472; AAA59063.1; -.
DR StGene: SG10191; Kbl.
DR InterPro: IPR001917; Aminotransf_2.
DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; PARTIAL.

```

KW Transferase; Acyltransferase; Pyridoxal phosphate.
 FT NON_TER 162 162
 SQ SEQUENCE 162 AA; 17671 MW; 9C1770AEB607392F CRC64;

Query Match 0.74; Score 7; DB 1; Length 162;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 210 GGLFETL 216
 |||||
 Db 116 GGLFETL 122

RESULT 27
 MLRS_MOUSE
 ID MLRS_MOUSE STANDARD; PRT; 168 AA.
 AC P97457;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN REGULATORY LIGHT CHAIN 2, SKELETAL MUSCLE ISOFORM (MLC2F).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Park K.W., Park W.J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96258423; PubMed=8777429;
 RA Paterno J., Gulick J., Ng W., Grupp I.L., Grupp G., Robbins J.;
 RT "Remodeling the mammalian heart using transgenesis."
 RL Cell Mol Biol Res 41:501-509(1995)
 CC -1- SUBUNIT: MYOSIN IS A HEXAMER OF 2 HEAVY CHAINS & 4 LIGHT CHAINS.
 CC -1- MISCELLANEOUS: THIS CHAIN BINDS CALCIUM.
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; U77943; AAB19118.1; -;
 DR HSSP; P02593; IAK8.
 DR SWISS-2DPAGE; P97457; MOUSE.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 2.
 DR PROSITE; PS00018; EF_HAND. 1.
 KW Myosin; Calcium-binding; Muscle protein; Phosphorylation.
 FT INT_MET 0
 FT MOD_RES 1 1 BY SIMILARITY.
 FT MOD_RES 15 15 BLOCKED (BY SIMILARITY).
 FT CA_BIND 37 48 PHOSPHORYLATION (BY SIMILARITY).
 FT CA_BIND 37 48 EF-HAND (POTENTIAL).
 SQ SEQUENCE 168 AA; 18824 MW; 03D401D66193377F CRC64;

Query Match 0.74; Score 7; DB 1; Length 168;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1014 VITGAFF 1020
 |||||
 Db 98 VITGAFF 104

RESULT 28

MLRS_RAT
 ID MLRS_RAT STANDARD; PRT; 168 AA.
 AC P04466;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN REGULATORY LIGHT CHAIN 2, SKELETAL MUSCLE ISOFORM (G2) (DTNB)
 DE (MLC-2).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85014159; PubMed=6091059;
 RA Nudel U., Calvo J.M., Shani M., Levy Z.;
 RT "The nucleotide sequence of a rat myosin light chain 2 gene."
 RL Nucleic Acids Res. 12:7175-7186(1984).
 RN [2]
 RP SEQUENCE OF 69-168 FROM N.A.
 RX MEDLINE=82265830; PubMed=6179945;
 RA Garfinkel L.I., Perlasamy M., Nadal-Ginard B.;
 RT "Cloning and characterization of cDNA sequences corresponding to
 RT myosin light chains 1, 2, and 3, troponin-C, troponin-T,
 RT alpha-tropomyosin, and alpha-actin."
 RL J. Biol. Chem. 257:11078-11086(1982).
 CC -1- SUBUNIT: MYOSIN IS A HEXAMER OF 2 HEAVY CHAINS & 4 LIGHT CHAINS.
 CC -1- MISCELLANEOUS: THIS CHAIN BINDS CALCIUM.
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; X00975; CA25480.1; -;
 DR EMBL; J00754; AAA41660.1; -;
 DR PIR; A03041; MORTL2.
 DR HSSP; P02593; IAK8.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 2.
 DR PROSITE; PS00018; EF_HAND. 1.
 KW Myosin; Calcium-binding; Muscle protein; Phosphorylation.
 FT INT_MET 0
 FT MOD_RES 1 1 BLOCKED (BY SIMILARITY).
 FT MOD_RES 15 15 PHOSPHORYLATION (BY SIMILARITY).
 FT CA_BIND 37 48 EF-HAND.
 SQ SEQUENCE 168 AA; 18838 MW; 27F05E7664696B45 CRC64;

Query Match 0.74; Score 7; DB 1; Length 168;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1014 VITGAFF 1020
 |||||
 Db 98 VITGAFF 104

RESULT 29
 MLRS_RABIT
 ID MLRS_RABIT STANDARD; PRT; 169 AA.
 AC P02608;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN REGULATORY LIGHT CHAIN 2, SKELETAL MUSCLE ISOFORM TYPE 2 (G2)
 DE (DTNB) (MLC-2).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91067462; PubMed=2147475;
RA Maeda K., Mueller-Gerhardt E., Wittinghofer A.;
RT "Sequence of two isoforms of myosin light chain 2 isolated from a
RL rabbit fast skeletal muscle lambda library.";
RN Nucleic Acids Res. 18:6687-6687(1990).
RN [2]
RP SEQUENCE OF 2-169.
RX MEDLINE=78216701; PubMed=352892;
RA Matsuda G., Malta T., Suzuyama Y., Setoguchi M., Umegane T.;
RT "The amino acid sequences of the tryptic, chymotryptic and peptic
RL peptides from the L-2 light chain of rabbit skeletal muscle myosin.";
RN Hoppe-Seyler's Z. Physiol. Chem. 359:629-640(1978).
RN [3]
RP SEQUENCE OF 2-169.
RX MEDLINE=77187770; PubMed=863872;
RA Matsuda G., Malta T., Suzuyama Y., Setoguchi M., Umegane T.;
RT "Amino acid sequence of the L-2 light chain of rabbit skeletal muscle
RL myosin.";
RN J. Biochem. 81:809-811(1977).
RN [4]
RP SEQUENCE OF 27-38 AND 112-145 FROM N.A.
RX MEDLINE=83167564; PubMed=6687628;
RA Putney S.D., Herlihy W.C., Schimmel P.R.;
RT "A new tropomyosin T and cDNA clones for 13 different muscle proteins,
RN found by shotgun sequencing.";
RL Nature 302:718-721(1983).
CC -1- SUBUNIT: MYOSIN IS A HEXAMER OF 2 HEAVY CHAINS & 4 LIGHT CHAINS.
CC -1- MISCELLANEOUS: THIS CHAIN BINDS ONE CALCIUM ION.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC DR EMBL: X54043; CAA37976.1; -
CC DR EMBL: V00887; CAA24255.1; -
CC DR EMBL: V00888; CAA24256.1; -
CC DR PIR: A03040; MORBD.
CC DR PIR: S12691; S12691.
CC DR PIR: S13445; S13445.
CC DR HSP: P02593; IAK8.
CC DR InterPro: IPR002048; EF-hand.
CC DR Pfam: PF00036; efhand; 2.
CC DR PROSITE: PS00018; EF_HAND; 1.
CC KM Myosin; Calcium-binding; Muscle protein; Phosphorylation.
CC FT INIT_MER 0 0
CC FT MOD_RES 1 1 BLOCKED.
CC FT MOD_RES 16 16 PHOSPHORYLATION (BY SIMILARITY).
CC FT CA_BIND 38 49 EF-HAND.
CC FT CA_BIND 38 49 EF-HAND.
CC SO SEQUENCE 169 AA; 18895 MW; AEDCA3BF1F9CA17A CRC64;

```

```

Query Match 0.7%; Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1014 VITGAFK 1020
DB 99 VITGAFK 105

```

```

RESULT 30
MLRT_RABIT STANDARD; PRT; 169 AA.
AC P24732;

```

```

DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE MYOSIN REGULATORY LIGHT CHAIN 2, SKELETAL MUSCLE ISOFORM TYPE 1 (G2)
DE (DTNB) (MLC-2).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91067462; PubMed=2147475;
RA Maeda K., Mueller-Gerhardt E., Wittinghofer A.;
RT "Sequence of two isoforms of myosin light chain 2 isolated from a
RL rabbit fast skeletal muscle lambda library.";
RN Nucleic Acids Res. 18:6687-6687(1990).
RN [2]
RP SEQUENCE OF 2-169.
RX MEDLINE=78216701; PubMed=352892;
RA Matsuda G., Malta T., Suzuyama Y., Setoguchi M., Umegane T.;
RT "The amino acid sequences of the tryptic, chymotryptic and peptic
RL peptides from the L-2 light chain of rabbit skeletal muscle myosin.";
RN Hoppe-Seyler's Z. Physiol. Chem. 359:629-640(1978).
RN [3]
RP SEQUENCE OF 2-169.
RX MEDLINE=77187770; PubMed=863872;
RA Matsuda G., Malta T., Suzuyama Y., Setoguchi M., Umegane T.;
RT "Amino acid sequence of the L-2 light chain of rabbit skeletal muscle
RL myosin.";
RN J. Biochem. 81:809-811(1977).
RN [4]
RP SEQUENCE OF 27-38 AND 112-145 FROM N.A.
RX MEDLINE=83167564; PubMed=6687628;
RA Putney S.D., Herlihy W.C., Schimmel P.R.;
RT "A new tropomyosin T and cDNA clones for 13 different muscle proteins,
RN found by shotgun sequencing.";
RL Nature 302:718-721(1983).
CC -1- SUBUNIT: MYOSIN IS A HEXAMER OF 2 HEAVY CHAINS & 4 LIGHT CHAINS.
CC -1- MISCELLANEOUS: THIS CHAIN BINDS ONE CALCIUM ION.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC DR EMBL: X54042; CAA37975.1; -
CC DR PIR: S12855; S12855.
CC DR HSP: P02593; IAK8.
CC DR InterPro: IPR002048; EF-hand.
CC DR Pfam: PF00036; efhand; 2.
CC DR PROSITE: PS00018; EF_HAND; 1.
CC KM Myosin; Calcium-binding; Muscle protein; Phosphorylation.
CC FT INIT_MER 0 0
CC FT MOD_RES 1 1 BLOCKED (BY SIMILARITY).
CC FT MOD_RES 16 16 PHOSPHORYLATION (BY SIMILARITY).
CC FT CA_BIND 38 49 EF-HAND.
CC FT CA_BIND 38 49 EF-HAND.
CC SO SEQUENCE 169 AA; 18897 MW; AEDP938C0A1689C3 CRC64;

```

```

Query Match 0.7%; Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1014 VITGAFK 1020
DB 99 VITGAFK 105

```

```

RESULT 31
Y786_METJA STANDARD; PRT; 186 AA.
AC Q58196;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE HYPOTHETICAL PROTEIN MJ0786.
GN MJ0786.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

```

RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.M., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.:
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*".
 RL Science 273:1058-1073(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67523; AAB98786.1; -
 DR TIGR: Mj0786; -
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 186 AA; 21628 MW; 769D3413D99977E7 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 186;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 IPDYLF 597
 |||||
 Db 140 IPDYLF 146

RESULT 32
 PH21_PSEFL STANDARD; PRT; 196 AA.
 AC Q51785;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AUTOINDUCER SYNTHESIS PROTEIN PH21.
 GN PH21.
 OS Pseudomonas fluorescens.
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 CC NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL B-15132 / 2-79;
 RX MEDLINE=98233753; PubMed=9573209;
 RA Mayrodi D.V., Ksenzenko V.N., Bonnell R.F., Cook R.J., Boronin A.M.,
 RA Thomashow L.S.;
 RT "A seven gene locus for synthesis of phenazine-1-carboxylic acid by
 RT *Pseudomonas fluorescens* 2-79.";
 RL J. Bacteriol. 180:2541-2548(1998).
 CC -I- FUNCTION: REQUIRED FOR THE SYNTHESIS OF A YET UNKNOWN N-ACETYL-
 CC HOMOSERINE LACTONE (N-ACETYL-HSL), AN AUTOINDUCER MOLECULE WHICH
 CC BINDS TO PHZ AND THUS REGULATES PHENAZINE PRODUCTION.
 CC -I- SIMILARITY: BELONGS TO THE AUTOINDUCER SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: LA8616; AAC18898.1; -
 DR InterPro: IPR001690; Autoinducers_synth.
 DR Pfam: PF00765; Autoind_synth. 1.
 DR PROSITE: PD002752; Autoinducers_synth. 1.
 DR PROSITE: PS00949; AUTOINDUCERS_SYNTH. 1.
 KW Quorum sensing; Autoinducer synthesis; Antibiotic biosynthesis.
 SO SEQUENCE 196 AA; 22399 MW; 24122CDAFA6E461 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 196;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 LECVFSH 404
 |||||
 Db 81 LECVFSH 87

RESULT 33
 YH13_YEAST STANDARD; PRT; 198 AA.
 ID YH13_YEAST
 AC P38896;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 21.1 KDA PROTEIN IN FLO5-PHO12 AND FLO1-PHO11 INTERGENIC
 DE REGIONS.
 GN YHR213W AND YAR062W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A. (YHR213W).
 RC STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favello A., Fulton L., Gattung S., Gelsel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (YAR062W).
 RC STRAIN=S288C / AB972;
 RA Bussey H., Keng T., Storms R.K., Vo D., Zhong W., Fortin N.,
 RA Barton A.B., Kaback D.B., Clark M.W.;
 RL Submitted (FEB-1994) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (YAR062W).
 RC STRAIN=S288C / AB972;
 RX MEDLINE=95249563; PubMed=7731988;
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
 RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
 RA Storms R.K.;
 RT "The nucleotide sequence of chromosome I from *Saccharomyces*
 RT *cerevisiae*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
 CC -I- SIMILARITY: TO YEAST PROTEIN FLO1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U00029; AAB69730.1; -
 DR EMBL: L28920; AAC09503.1; -
 DR PIR: S48994; S48994.
 DR InterPro: IPR001389; Flocculin.
 DR Pfam: PF00624; Flocculin. 1.
 KW Hypothetical protein.
 FT DOMAIN 133 198 SER/THR-RICH.
 SO SEQUENCE 198 AA; 21145 MW; 49B50117AEAF7DE5 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 198;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 IVTTTTE 282
 |||||
 Db 136 IVTTTTE 142

RESULT 34
 OPBD_BACSU STANDARD; PRT; 226 AA.
 AC P39775: Q34657:
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHOLINE TRANSPORT SYSTEM PERMEASE PROTEIN OPUBD.
 GN OPUBD OR PROZ.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 6633 / LH45;
 RX MEDLINE=96074318; PubMed=7592481;
 RA Lin Y., Hansen J.N.;
 RT "Characterization of a chimeric proU operon in a subtilin-producing
 mutant of Bacillus subtilis 168.";
 RL J. Bacteriol. 177:6874-6880(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RX MEDLINE=9932519; PubMed=10216873;
 RA Kappes R.M., Kempf B., Kneip S., Boch J., Gade J., Meier-Wagner J.,
 RA Bremer E.;
 RT "Two evolutionarily closely related ABC transporters mediate the
 uptake of choline for synthesis of the osmoprotectant glycine betaine
 in Bacillus subtilis.";
 RL Mol. Microbiol. 32:203-216(1999).
 RN [3]
 RP SEQUENCE OF 21-226 FROM N.A.
 RC STRAIN-ATCC 6633 / LH45;
 RX MEDLINE=93015727; PubMed=1400221;
 RA Chung Y.J., Hansen J.N.;
 RT "Determination of the sequence of spaE and identification of a
 promoter in the subtilin (spa) operon in Bacillus subtilis.";
 RL J. Bacteriol. 174:6699-6705(1992).
 CC -1- FUNCTION: INVOLVED IN A HIGH AFFINITY MULTICOMPONENT BINDING-
 PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR CHOLINE. PROBABLY
 RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE
 MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYSTM
 SUBFAMILY
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U38418; AAB01535.1; -
 CC EMBL; AF008930; AAC14359.1; -
 CC EMBL; M99263; AAA22773.1; -
 CC EMBL; Z99121; CAB15375.1; -
 CC SdbLisc; BG12636; OPUBD.
 CC InterPro: IPR000515; BPD_transp.
 CC Pfam: PF00528; BPD_transp; 1.

DR PROSITE; PS00402; BPD_TRANS_PNN_MEMBR; 1.
 KW Transmem; Amino-acid transport; Transmembrane; Complete proteome.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 183 203 POTENTIAL.
 FT VARIANT 30 30 A -> V (IN STRAIN LH45).
 FT VARIANT 43 43 F -> Y (IN STRAIN LH45).
 FT VARIANT 50 50 L -> I (IN STRAIN LH45).
 FT VARIANT 89 89 I -> V (IN STRAIN LH45).
 FT VARIANT 108 108 I -> T (IN STRAIN LH45).
 FT VARIANT 189 189 V -> I (IN STRAIN LH45).
 FT VARIANT 196 196 L -> I (IN STRAIN LH45).
 FT VARIANT 201 201 L -> I (IN STRAIN LH45).
 FT VARIANT 205 205 L -> I (IN STRAIN LH45).
 FT CONFLICT 208 210 ALS -> FLN (IN REF. 1 AND 3).
 FT CONFLICT 214 226 KKRTGAKHVQSA -> QKRRRIVSV (IN REF. 1
 AND 3).
 SQ SEQUENCE 226 AA; 23927 MW; DA6D47BAF68347E9 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 226;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 875 MNVLEQL 881
 |||||
 Db 1 MNVLEQL 7

RESULT 35
 UCRL_SOLITU STANDARD; PRT; 265 AA.
 ID UCRL_SOLITU
 AC P37841;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE UBIOJINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT, MITOCHONDRIAL
 DE PRECURSOR (EC 1.10.2.2) (RIESKE IRON-SULFUR PROTEIN) (RISP).
 GN FEBS1.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=CV. DESIREE; TISSUE=Leaf;
 RX MEDLINE=94289650; PubMed=8018875;
 RA Emmertmann M., Clericus M., Braun H.P., Moxo T., Helms L., Kruff V.,
 RA Schmitz U.K.;
 RT "Molecular features, processing and import of the Rieske iron-sulfur
 RT protein from potato mitochondria.";
 RL Plant Mol. Biol. 25:271-281(1994).
 CC -1- FUNCTION: COMPONENT OF THE UBIOJINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +
 CC 2 FERROCYTOCHROME C.
 CC -1- SUBUNIT: BCI COMPLEX CONTAINS 10 SUBUNITS, 3 RESPIRATORY
 CC SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
 CC -1- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S
 CC PROTEIN.
 CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
 CC BACTERIAL, CHLOROPLAST).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: X79332; CAA55894.1; -
 DR PIR: S45037; S45037.
 DR HSSP: P13272; IRIE.
 DR InterPro: IPR001281; Rieske.
 DR Pfam: PF00355; Rieske; 1.
 DR PRINTS: PR00162; RIESKE.
 DR PROSITE: PS00199; RIESKE_1; 1.
 DR PROSITE: PS00200; RIESKE_2; 1.
 DR Mitochondrion: Electron transport; Respiratory chain; Iron-sulfur;
 KW Oxidoreductase; Inner membrane; Transmembrane; Transit peptide.
 FT TRANSIT 1 53
 FT CHAIN 54 265
 FT METAL 208 208
 FT METAL 210 210
 FT METAL 227 227
 FT METAL 230 230
 FT DISULFID 213 229
 SQ SEQUENCE 265 AA; 28840 MW; 78C8C0AC6F09D932 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 265;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 451 AGRRLSS 457
 |||||
 Db 5 AGRRLSS 11

RESULT 36

UCRL_MAIZE
 ID UCRL_MAIZE STANDARD; PRT: 273 AA.

AC P49727;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT, MITOCHONDRIAL
 DE PRECURSOR (EC 1.10.2.2) (RIESKE IRON-SULFUR PROTEIN) (RISP).
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Palcoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;

RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=92073358; PubMed=1961737;

RA Huang J.T., Struck F., Matzinger D.F., Levings C.S. III;

"Functional analysis in yeast of cDNA coding for the mitochondrial

Rieske iron-sulfur protein of higher plants.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:10716-10720(1991).

-1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A

RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

COUPLED TO ATP SYNTHESIS.

-1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = Q +

2 FERROCYTOCHROME C.

-1- SUBUNIT: BCI COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY

SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.

-1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.

-1- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S

PROTEIN.

-1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,

BACTERIAL, CHLOROPLAST).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: M7724; AAA3507.1; -
 DR HSSP: P13272; IRIE.
 DR MaizeDB: 30151; -
 DR InterPro: IPR001281; Rieske.
 DR Pfam: PF00355; Rieske; 1.
 DR PRINTS: PR00162; RIESKE.
 DR PROSITE: PS00199; RIESKE_1; 1.
 DR PROSITE: PS00200; RIESKE_2; 1.
 DR Mitochondrion: Electron transport; Respiratory chain; Iron-sulfur;
 KW Oxidoreductase; Inner membrane; Transmembrane; Transit peptide.
 FT TRANSIT 1 61
 FT CHAIN 62 273
 FT METAL 216 216
 FT METAL 218 218
 FT METAL 235 235
 FT METAL 238 238
 FT DISULFID 221 237
 SQ SEQUENCE 273 AA; 29835 MW; 43B08C116E9B1A2C CRC64;

Query Match 0.7%; Score 7; DB 1; Length 273;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 451 AGRRLSS 457
 |||||
 Db 5 AGRRLSS 11

RESULT 37

POR3_BOVIN
 ID POR3_BOVIN STANDARD; PRT: 283 AA.

AC O9W213;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (HVDAC3)
 DE (OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
 GN VDAC3.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;

RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Lens;

RA Rae J.L.;

"Ion channels in the lens.";

RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.

-1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER

MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES

(BY SIMILARITY).

-1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.

-1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.

-1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: AF268466; AAF80103.1; -
 CC InterPro: IPR001925; Euk_porin.
 CC Pfam: PF01459; Euk_porin; 1.

DR PRINTS: PR00185; EUKARYOTIC.PORIN.
 DR PROSITE: PS00558; EUKARYOTIC.PORIN: 1.
 KW Outer membrane; Porin; Mitochondrion..
 SQ SEQUENCE 283 AA; 30739 MW; D30SDA2EE42BEC3B CRC64;

Query Match 0.7%: Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 782 KLAEGLK 788
 |||||
 Db 90 KLAEGLK 96

RESULT 38
 ID POR3_HUMAN STANDARD; PRT; 283 AA.
 AC Q9Y277; Q9UIS0;
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (HVDAC3)
 DE (OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
 GN VDAC3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=98318631; PubMed=9653160;
 RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
 He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
 Wang Y.-X., Chen S.-J., Chen Z.;
 RT "Identification of genes expressed in human CD34(+) hematopoietic
 RT stem/progenitor cells by expressed sequence tags and efficient full-
 RT length cDNA cloning.";
 RT Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98454325; PubMed=9781040;
 RA Rahmani Z., Maunoury C., Siddiqui A.;
 RT "Isolation of a novel human voltage-dependent anion channel gene.";
 RL Eur. J. Hum. Genet. 6:337-340(1998).
 RN [3]
 RP SEQUENCE OF 1-253 FROM N.A.
 RX MEDLINE=99431679; PubMed=10501981;
 RA Decker W.K., Bowles K.R., Schatte E.C., Towbin J.A., Craigen W.J.;
 RT "Revised fine mapping of the human voltage-dependent anion channel
 RT loci by radiation hybrid analysis.";
 RL Mamm. Genome 10:1041-1042(1999).
 RN [4]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=20295349; PubMed=10833333;
 RA Decker W.K., Craigen W.J.;
 RT "The tissue-specific, alternatively spliced single ATG exon of the
 RT type 3 voltage-dependent anion channel gene does not create a
 RT truncated protein isoform in vivo.";
 RL Mol. Genet. Metab. 70:69-74(2000).
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
 CC MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, HIGHEST IN TESTIS.
 CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF038962; AAC39876.1; -
 DR EMBL: U90943; AAB93872.1; -
 DR EMBL: AF151682; AAD49610.1; -
 DR EMBL: AF151679; AAD49610.1; JOINED.
 DR EMBL: AF151681; AAD49610.1; JOINED.
 DR InterPro: IPR001925; Euk_porin.
 DR Pfam: PF01459; Euk_porin; 1.
 DR PRINTS: PR00185; EUKARYOTIC.PORIN.
 DR PROSITE: PS00558; EUKARYOTIC.PORIN: 1.
 KW Outer membrane; Porin; Mitochondrion; Alternative splicing.
 FT VARSPPLIC 39 V -> VM (IN ISOFORM 2).
 SQ SEQUENCE 283 AA; 30658 MW; E03CBCEDA72A9783 CRC64;

Query Match 0.7%: Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 782 KLAEGLK 788
 |||||
 Db 90 KLAEGLK 96

RESULT 39
 ID POR3_MOUSE STANDARD; PRT; 283 AA.
 AC Q60931;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (MVDAC3)
 DE (OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
 GN VDAC3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=96411667; PubMed=8812436;
 RA Sampson M.J., Lovell R.S., Davison D.B., Craigen W.J.;
 RT "A novel mouse mitochondrial voltage-dependent anion channel gene
 RT localizes to chromosome 8.";
 RL Genomics 36:192-196(1996).
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
 CC MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION DETECTED IN
 CC TESTIS, LESS BUT STILL ABUNDANT EXPRESSION IN HEART, KIDNEY,
 CC BRAIN, AND SKELETAL MUSCLE.
 CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U30839; AAB4776.1; -
 DR MGD: MGI:106922; Vdac3.
 DR InterPro: IPR001925; Euk_porin.
 DR Pfam: PF01459; Euk_porin; 1.

DR PRINTS; PRO0185; EUKARYOTIC.PORIN.1.
 DR PROSITE; PS00558; EUKARYOTIC.PORIN.1.
 KW Outer membrane; Porin; Mitochondrion.
 SQ SEQUENCE 283 AA; 30753 MW; A95AFD67C611228C CRC64;

Query Match 0.78; Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGLK 788
 |||||
 DB 90 KLAEGLK 96

RESULT 40
 POR3_RABIT STANDARD; PRT; 283 AA.

AC O9T13;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (OUTER
 DE MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
 GN VDAC3.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_Taxid:9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEW ZEALAND WHITE; TISSUE=Cornea;
 RA Rae J.L.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
 CC MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
 CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AF209727; AAF22837.1; -
 CC InterPro; IPR001925; Euk_porin.
 CC Pfam; PF01459; Euk_porin.1.
 DR PRINTS; PRO0185; EUKARYOTIC.PORIN.
 DR PROSITE; PS00558; EUKARYOTIC.PORIN.1.
 KW Outer membrane; Porin; Mitochondrion.
 SQ SEQUENCE 283 AA; 30651 MW; BCC05616366090A0 CRC64;

Query Match 0.78; Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGLK 788
 |||||
 DB 90 KLAEGLK 96

RESULT 41
 POR3_RAT STANDARD; PRT; 283 AA.
 AC O9R1Z0; O9WTU2; O9ESR2; O9J131;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (RVDAC3)
 DE (OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
 GN VDAC3.
 OS Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid:10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20453129; PubMed=10998068;
 RA Shinozawa Y., Ishida T., Hino M., Yamazaki N., Baba Y., Terada H.;
 RT "Characterization of porin isoforms expressed in tumor cells";
 RL Eur. J. Biochem. 267:6067-6073(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=Sprague-Dawley;
 RA Rae J.L.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RL [3]
 RP SEQUENCE OF 10-283 FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Heart;
 RX MEDLINE=98390268; PubMed=9714728;
 RA Anflous K., Blondel O., Bernard A., Khrestchatsky M.,
 RA Ventura-Clapier R.;
 RT "Characterization of rat porin isoforms: cloning of a cardiac type-3
 RT variant encoding an additional methionine at its putative N-terminal
 RT region";
 RL Blochim. Biophys. Acta 1399:47-50(1998).
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
 CC MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1/RVDAC3 (SHOWN HERE) AND
 CC 2/RVDAC3V; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: RVDAC3 IS WIDELY EXPRESSED, STRONGEST IN
 CC ATRIA. RVDAC3V IS ALSO WIDELY EXPRESSED, HIGHEST IN BRAIN BUT NOT
 CC EXPRESSED IN KIDNEY.
 CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AB039664; BAB13475.1; -
 CC EMBL; AF268469; AAF80117.1; -
 CC EMBL; AF048829; AAD22722.1; -
 CC EMBL; AF048830; AAD22723.1; -
 CC InterPro; IPR001925; Euk_porin.
 CC Pfam; PF01459; Euk_porin.1.
 DR PROSITE; PS00558; EUKARYOTIC.PORIN.1.
 KW Outer membrane; Porin; Mitochondrion; Alternative splicing.
 KW VARSPLIC
 FT CONFLICT 39 128 V -> VM (IN ISOFORM 2).
 FT CONFLICT 128 128 K -> N (IN REF. 2).
 SQ SEQUENCE 283 AA; 30798 MW; 38002466b557864 CRC64;

Query Match 0.78; Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGLK 788
 |||||
 DB 90 KLAEGLK 96

RESULT 42

Y388 AQUAE STANDARD: PRT; 301 AA.
 ID Y388 AQUAE
 AC 066709;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOHETICAL PROTEIN AQ_388.
 GN AQ_388.
 OS Aquifex aeolicus.
 OC Bacteria: Aquificales: Aquificaceae: Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus";
 RL Nature 392:353-358(1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AE000687; AAC06669.1; -
 CC DR Hypothetical protein: Complete proteome.
 SQ SEQUENCE 301 AA; 35030 MW; C7E22EDCE8017872 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 301;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 786 GLKRLKK 792
 |||||
 DB 120 GLKRLKK 126

RESULT 43
 FMT_HELPJ STANDARD: PRT; 305 AA.
 ID FMT_HELPJ
 AC 092K72;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE METHIONYL-TRNA FORMYLTRANSFERASE (EC 2.1.2.9).
 GN FMT OR JHP1069.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria: Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori";
 RL Nature 397:176-180(1999).
 CC -----
 CC -!- FUNCTION: MODIFY THE FREE AMINO GROUP OF THE AMINOACYL MOIETY OF
 CC METHIONYL-TRNA(FMT). THE FORMYL GROUP APPEARS TO PLAY A DUAL ROLE
 CC IN THE INITIATOR IDENTITY OF N-FORMYLMETHIONYL-TRNA BY: (1)
 CC PROMOTING ITS RECOGNITION BY IF2 AND (II) IMPAIRING ITS BINDING TO
 CC EFU-GTP (BY SIMILARITY).
 CC -----

CC -!- CATALYTIC ACTIVITY: 10-FORMYLTETRAHYDROFOLATE + L-METHIONYL-TRNA
 CC + H(2)O = TETRAHYDROFOLATE + N-FORMYLMETHIONYL-TRNA.
 CC -!- SIMILARITY: BELONGS TO THE FMT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AE001534; AAD06649.1; -
 CC DR HSSP: P23882; 1FMT.
 DR InterPro: IPR001555; GART.
 DR InterPro: IPR002376; formyl_transf.
 DR Pfam: PF00551; formyl_transf; 1.
 DR PROSITE: PS00373; GART; FALSE_NEG.
 KM Transferrase: Methyltransferase; Protein biosynthesis;
 KM Complete proteome.
 FT BINDING 111
 FT SEQUENCE 305 AA; 34233 MW; 498E7C1F60A42D8 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 305;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 67 NLEFLSL 73
 |||||
 DB 218 NLEFLSL 224

RESULT 44
 ABCA_AERSA STANDARD: PRT; 308 AA.
 ID ABCA_AERSA
 AC 007698;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ABC TRANSPORTER PROTEIN ABCA.
 GN ABCA.
 OS Aeromonas salmonicida.
 OC Bacteria: Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OX NCBI_TaxID=645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A450;
 RX MEDLINE=9325958; PubMed=8491726;
 RA Chu S., Trust T.J.;
 RT "An Aeromonas salmonicida gene which influences a-protein expression
 RT in Escherichia coli encodes a protein containing an ATP-binding
 RT cassette and maps beside the surface array protein gene.";
 RL J. Bacteriol. 175:3105-3114(1993).
 CC -----
 CC -!- FUNCTION: INFLUENCES THE EXPRESSION OF THE SURFACE ARRAY PROTEIN
 CC GENE (VAP). MAY HAVE BOTH REGULATORY AND TRANSPORT ACTIVITIES.
 CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: L11870; AAA21933.1; -
 DR PTR: A36918; A36918.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR001687; ATP_GTP_A.

DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding. 58 65 ATP (POTENTIAL).
 FT NP_BIND 238 259 ARG/LYS-RICH (BASIC).
 FT DOMAIN 308 AA; 34015 MW; 412023797d883e18 CRC64;
 SQ SEQUENCE

Query Match 0.7%; Score 7; DB 1; Length 308;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180
 |||||
 Db 63 GKSTLLQ 69

RESULT 45
 CWL1_SCHPO STANDARD; PRT; 308 AA.

AC P53694;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CELL LYSIS PROTEIN CWL1.
 GN CWL1 OR SPBC31A8.01C OR SPBC651.13C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 SEQUENCE FROM N.A.

RA MEDLINE=97027308; PubMed=8873452;
 RX Godoy C., Arellano M., Diaz M., Duran A., Perez P.;
 RT "Characterization of cwl1+, a gene from Schizosaccharomyces pombe
 RL whose overexpression causes cell lysis.";
 RL Yeast 12:983-990(1996).
 RN [2]
 SEQUENCE FROM N.A.

RP STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- OVEREXPRESSION CAUSES CELL LYSIS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X94445; CAAG4219.1; -
 DR EMBL; AL035384; CAB37626.1; -
 DR EMBL; AL035570; CAB37609.1; -
 KW Cell wall.
 SQ SEQUENCE 308 AA; 33659 MW; F6FA92A6E9F695D CRC64;

Query Match 0.7%; Score 7; DB 1; Length 308;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LFVWNVL 33
 |||||
 Db 258 LFVWNVL 264

RESULT 46
 ASFG_BACLI

ID ASFG_BACLI STANDARD; PRT; 322 AA.
 AC P30363;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE L-ASPARAGINASE (EC 3.5.1.1) (L-ASPARAGINE AMIDOHYDROLASE) (L-ASNASE).
 GN ANSA.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 SEQUENCE FROM N.A.

RA MEDLINE=92009126; PubMed=1916233;
 RX van Dijk J.M., de Jong A., Bron S., Venema G.;
 RT "Lack of specific hybridization between the lep genes of Salmonella
 RT typhimurium and Bacillus licheniformis.";
 RL FEMS Microbiol. Lett. 65:345-351(1991).
 CC -1- CATALYTIC ACTIVITY: L-ASPARAGINE + H(2)O = L-ASPARTATE + NH(3).
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE ASPARAGINASE 1 FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z11497; CAAT7574.1; -
 DR PIR; S18999; S18999.
 DR InterPro; IPR000267; Asparaginase-glutaminase.
 DR Pfam; PF00710; Asparaginase; 1.
 DR PRINTS; PR00139; ASNGINASE.
 DR PROSITE; PS00144; ASN_GLN_ASE_1; 1.
 DR PROSITE; PS00917; ASN_GLN_ASE_2; 1.
 DR Hydrolyase.
 FT ACT_SITE 13 13 BY SIMILARITY.
 FT ACT_SITE 89 89 BY SIMILARITY.
 FT ACT_SITE 90 90 BY SIMILARITY.
 FT ACT_SITE 162 162 BY SIMILARITY.
 SQ SEQUENCE 322 AA; 35442 MW; 7048BFC520B1DFB CRC64;

Query Match 0.7%; Score 7; DB 1; Length 322;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 724 LTIEDER 730
 |||||
 Db 101 LTIEDER 107

RESULT 47
 OSTY_YEAST

AC 003723;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE DOLICHYL-DIPHOSPHOLIGOSACCHARIDE--PROTEIN GLYCOSYLTRANSFERASE 37 KDA
 DE SUBUNIT (EC 2.4.1.119) (OLIGOSACCHARYL TRANSFERASE 37 KDA SUBUNIT)
 DE (ORFase 37 KDA SUBUNIT).
 GN OST6 OR YML019W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Krauer R.;

RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
 Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: SUBUNIT OF OLIGOSACCHARYL TRANSFERASE ENZYME WHICH
 CC CATALYSES THE TRANSFER OF A HIGH MANNOSE OLIGOSACCHARIDE TO AN
 CC ASPARAGINE RESIDUE WITHIN AN ASN-X-SER/THR CONSENSUS MOTIF IN
 CC NASCENT POLYPEPTIDE CHAINS.
 CC -1- CATALYTIC ACTIVITY: DOLICHYL DIPHOSPHOOLIGOSACCHARIDE + PROTEIN
 CC L-ASPARAGINE = DOLICHYL DIPHOSPHATE + A GLYCOPROTEIN WITH THE
 CC OLIGOSACCHARIDE CHAIN ATTACHED BY GLYCOSYLAMINE LINKAGE TO PROTEIN
 CC L-ASPARAGINE.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y08606; CAA69898.1; -
 DR EMBL: Z46659; CAA66636.1; -
 DR SGD: S0004481; OST6.
 DR Transferase: Transmembrane.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 268 288 POTENTIAL.
 FT TRANSMEM 303 323 POTENTIAL.
 SQ SEQUENCE 332 AA; 37890 MW; FC7AE156BF27F1B5 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 332;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 905 EYPOLVK 911
 |||||
 Db 113 EYPOLVK 119

RESULT 48
 ADH_SULSR
 ID ADH_SULSR STANDARD; PRT; 347 AA.
 AC P50381;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NAD-DEPENDENT ALCOHOL DEHYDROGENASE (EC 1.1.1.1).
 GN ADH.
 OS Sulfolobus sp. (strain RC3).
 OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=165757;
 RX MEDLINE=96125263; PubMed=8550434;
 RA Canino R., Fiorentino G., Carpinelli P., Rossi M., Bartolucci S.;
 RT "Cloning and overexpression in Escherichia coli of the genes encoding
 RT NAD-dependent alcohol dehydrogenase from two Sulfolobus species.";
 RL J. Bacteriol. 178:301-305(1996).
 CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
 CC -1- COFACTOR: BINDS FOUR ZINC IONS PER DIMER.
 CC -1- SUBUNIT: HOMODIMER AND HOMOTETRAMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z47543; CAA87591.1; -
 DR HSSP: P00325; 1HDY.
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zinc.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; zinc; NAD; Methylation.
 FT MOD_RES 11 11 METHYLATION (BY SIMILARITY).
 FT MOD_RES 213 213 METHYLATION (BY SIMILARITY).
 FT METAL 38 38 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 68 68 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 101 101 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 112 112 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 347 AA; 37583 MW; 55D3C003792DE60D CRC64;

Query Match 0.7%; Score 7; DB 1; Length 347;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 719 VEASPLT 725
 |||||
 Db 147 VEASPLT 153

RESULT 49
 YCFE_ECOLI
 ID YCFE_ECOLI STANDARD; PRT; 357 AA.
 AC P75955;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 40.8 KDA PROTEIN IN MFD-COBB INTERGENIC REGION.
 GN YCFE OR B1115.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Ikeda T., Itoh T., Kajiya M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AE000212; AAC74199.1; -
 DR EMBL: D90746; BAA35930.1; -
 DR EMBL: D90747; BAA35935.1; -
 DR EcGene: EG13438; ycfT
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 136 156 POTENTIAL.
 FT TRANSMEM 163 183 POTENTIAL.
 FT TRANSMEM 197 217 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 SQ SEQUENCE 357 AA; 40833 MW; F1ED5FBC7781A4F9 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 357;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 LVKRLSQ 989
 |||||
 DB 330 LVKRLSQ 336

RESULT 50
 RFL_CHLMU STANDARD; PRT; 359 AA.
 AC 09PIL6;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PEPTIDE CHAIN RELEASE FACTOR 1 (RF-1).
 GN PREP OR TC0292.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / NIGG;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- FUNCTION: PEPTIDE CHAIN RELEASE FACTOR 1 DIRECTS THE TERMINATION
 CC OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN TERMINATION
 CC CODONS UAG AND UAA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
 CC FACTORS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AE002297; AAF39160.1; -
 DR TIGR; TIGR0292; -

DR InterPro: IPR000352; Pep_rel_factor-I.
 DR Pfam: PF00472; RF-1; 1.
 DR PROSITE: PS00745; RF_PROK-I; 1.
 KW Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 359 AA; 40207 MW; 2F58985DA434038B CRC64;

Query Match 0.7%; Score 7; DB 1; Length 359;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1009 DDDLSTVI 1015
 |||||
 DB 105 DDDLSTVI 111

RESULT 51
 RFL_CHLTR STANDARD; PRT; 359 AA.
 AC 084026;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PEPTIDE CHAIN RELEASE FACTOR 1 (RF-1).
 GN PREP OR CT023.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/WR-3/CX;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusev R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- FUNCTION: PEPTIDE CHAIN RELEASE FACTOR 1 DIRECTS THE TERMINATION
 CC OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN TERMINATION
 CC CODONS UAG AND UAA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
 CC FACTORS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AE001277; AAC67613.1; -
 DR InterPro: IPR000352; Pep_rel_factor-I.
 DR Pfam: PF00472; RF-1; 1.
 DR PROSITE: PS00745; RF_PROK-I; 1.
 KW Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 359 AA; 40052 MW; 0ED087BCC230CE7A CRC64;

Query Match 0.7%; Score 7; DB 1; Length 359;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1009 DDDLSTVI 1015
 |||||
 DB 105 DDDLSTVI 111

RESULT 52
 YB19_SYNY3 STANDARD; PRT; 361 AA.
 ID YB19_SYNY3

AC P73341;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 40.9 KDA PROTEIN SL1119.
GN SL1119.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugijura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996)
CC -1- SIMILARITY: BELONGS TO THE PSEUDOMONAS-TYPE THRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90905; BAA17372.1; -
CC DR Hypothetical protein: Complete proteome.
CC KW SEQUENCE 361 AA; 40913 MW; 25C2933F9F8205A5 CRC64;
SQ

Query Match 0.7%; Score 7; DB 1; Length 361;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 431 TAQRFKP 437
IIIIII
DB 175 TAQRFKP 181
RESULT 53
HRC_A_SYN3
ID HRC_A_SYN3 STANDARD; PRT; 378 AA.
AC P72795;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HEAT-INDUCIBLE TRANSCRIPTION REPRESSOR HRC_A_HOMOLOG.
GN HRC_A_SIL1670.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugijura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: NEGATIVE REGULATOR OF CLASS I HEAT SHOCK GENES (GRPE-
CC DNA-DNAU AND GROELS OPERONS). PREVENTS HEAT-SHOCK INDUCTION OF
CC THESE OPERONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HRC_A FAMILY.
CC -----

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90900; BAA16810.1; -
CC DR InterPro: IPR002571; HRC_A.
DR Pfam: PF01628; HRC_A; 1.
KW Transcription regulation; Repressor; Heat shock; Complete proteome.
SQ SEQUENCE 378 AA; 42626 MW; DBB24C059ED04AFBC CRC64;
SQ

Query Match 0.7%; Score 7; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ALIORMG 16
IIIIII
DB 111 ALIORMG 117

RESULT 54
XYLR_STAXY
ID XYLR_STAXY STANDARD; PRT; 383 AA.
AC P27159;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE XYLOSE REPRESSOR.
GN XYLR.
OS Staphylococcus xylosus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group.
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1288;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20267 / C2A;
RX MEDLINE=9136026; PubMed=1714034;
RA Sizemore C., Buchner E., Rygus T., Witke C., Goetz F., Hillen W.;
RT "Organization, promoter analysis and transcriptional regulation of
RT the Staphylococcus xylosus xylose utilization operon.";
RL Mol. Gen. Genet. 227:377-384(1991).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF XYLOSE-UTILIZING ENZYMES.
CC -1- SIMILARITY: BELONGS TO THE ROK (NAC/XYLR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57599; CAA40823.1; -
CC DR PIR: S16529; S16529.
DR InterPro: IPR000600; ROK.
DR Pfam: PF00480; ROK; 1.
DR PROSITE: PS01125; ROK; 1.
KW Transcription regulation; DNA-binding; Repressor; Xylose metabolism.
FT DNA_BIND 26 45
SQ SEQUENCE 383 AA; 43376 MW; 601CCD3D3D64ECAF CRC64;
SQ

Query Match 0.7%; Score 7; DB 1; Length 383;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 582 KSLVINS 588
IIIIII

```

Db      194 KSIXINS 200

RESULT  55
Y464_MYCPN
ID      Y464_MYCPN      STANDARD;      PRT;      365 AA.
AC      P75112;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HYPOTHETICAL PROTEIN MG464 HOMOLOG (K05_ORF385).
GN      MEN680 OR MP162.
OS      Mycoplasma pneumoniae.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC      Mycoplasmataceae; Mycoplasma.
OX      NCBI_TaxID=2104;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 29342 / M129;
RX      MEDLINE=97105885; PubMed=8948633;
RA      Himmelfreisch R., Hilbert H., Plagens H., Li B.-C.,
RA      Herrmann R.;
RT      "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT      pneumoniae.";
RL      Nucleic Acids Res. 24:4420-4449(1996).
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC      -1- SIMILARITY: TO A M.CAPRICOLUM HOMOLOG.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE000017; AAB95810.1; -
DR      KW Hypothetical protein; Transmembrane; Complete proteome.
FT      TRANSMEM 39 59 POTENTIAL.
FT      TRANSMEM 150 170 POTENTIAL.
FT      TRANSMEM 222 242 POTENTIAL.
FT      TRANSMEM 270 290 POTENTIAL.
FT      TRANSMEM 325 345 POTENTIAL.
FT      SEQUENCE 385 AA; 44239 MW; C85B5A28DC5B82D CRC64;
SO      SEQUENCE

Query Match      0.7%; Score 7; DB 1; Length 385;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      786 GLKLNKK 792
      |||||
Db      30 GLKLNKK 36

RESULT  56
SYM_AQUAE
ID      SYM_AQUAE      STANDARD;      PRT;      395 AA.
AC      O67115;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      TRYPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2) (TRYPTOPHAN--TRNA LIGASE)
DE      (TRPRS).
GN      TRPS OR AQ.992.
OS      Aquifex aeolicus.
OC      Bacteria; Aquificales; Aquificaceae; Aquifex.
OX      NCBI_TaxID=63363;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-VF5;
RX      MEDLINE=98196666; PubMed=9537320;
RA      Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

```

```

RA      Graham D.E., Overbeek R., Sneed M.A., Keller M., Anujay M., Huber R.,
RA      Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT      "The complete genome of the hyperthermophilic bacterium Aquifex
RT      aeolicus.";
RL      Nature 392:353-358(1998).
CC      -1- CATALYTIC ACTIVITY: ATP + L-TRYPTOPHAN + TRNA(Trp) = AMP +
CC      PYROPHOSPHATE + L-TRYPTOPHANYL-TRNA (TRP).
CC      -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE000717; AAC07073.1; -
DR      InterPro: IPR002305; tRNA-synt_1b.
DR      InterPro: IPR001412; tRNA-synt_1.
DR      InterPro: IPR002306; tRNA-synt_trp.
DR      Pfam: PF00579; tRNA-synt_1b; 1.
DR      PRINTS: PR01039; TRNASYNTTRP.
DR      PROSITE: PS00178; AA-TRNA_LIGASE_I; FALSE_NEG.
KW      Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW      Complete proteome.
FT      SITE 9 17 "HIGH" REGION.
FT      BINDING 261 265 "KMSKS" REGION.
FT      BINDING 264 264 ATP (BY SIMILARITY).
FT      SEQUENCE 395 AA; 46709 MW; 72345F53CCEFB8F5 CRC64;
SO      SEQUENCE

Query Match      0.7%; Score 7; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      899 LKHLKE 905
      |||||
Db      341 LKHLKE 347

RESULT  57
PRRC_ECOLI
ID      PRRC_ECOLI      STANDARD;      PRT;      396 AA.
AC      P17223;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      01-FEB-1995 (Rel. 31, Last annotation update)
DE      ANTICODON NUCLEASE.
DE      PRRC.
GN      Escherichia coli.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CTR5X;
RX      MEDLINE=90228334; PubMed=1691706;
RA      Levitz R., Chapman D., Amitsur M., Green R., Snyder L., Kaufmann G.;
RT      "The optional E. coli p1r locus encodes a latent form of phage T4-
RT      induced anticodon nuclease.";
RL      EMBO J. 9:1383-1389(1990).
CC      -1- FUNCTION: ANTICODON NUCLEASE TRIGGERS THE CLEAVAGE LIGATION OF
CC      TRNA(TXS). IT IS ACTIVATED BY T4 STP PROTEIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```


CC or send an email to license@lsb-sib.ch.
 CC EMBL: X52284; CAA36527.1; -
 DR PIR: S09627; S09627.
 KW Hydrolyase; Nuclease.
 SO SEQUENCE 396 AA; 45568 MW; 607C9C9BC99C0894 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 499 VENTRAV 505
 |||||
 DB 26 VENTRAV 32

RESULT 58
 KBL_ECOLI STANDARD; PRT; 398 AA.
 AC P07912;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 2-AMINO-3-KETOBUTYRATE COENZYME A LIGASE (EC 2.3.1.29) (AKB LIGASE)
 DE (GLYCINE ACETYLTRANSFERASE).
 OS KBL OR B3617 OR Z5044 OR ECS4495.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA Amonson B.D., Ravnikar P.D., Somerville R.L.;
 RT "Nucleotide sequence of the 2-amino-3-ketobutyrate coenzyme A ligase
 (kbl) gene of E. coli".
 RL Nucleic Acids Res. 16:3586-3586(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 region from 76.0 to 81.5 minutes".
 RL Nucleic Acids Res. 22:2576-2586(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7".
 RL Nature 409:529-533(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 050952;
 MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12".
 RL DNA Res. 8:11-22(2001).
 RN [5]
 RP SEQUENCE OF 1-21 FROM N.A.

RX MEDLINE=88032988; PubMed=3117785;
 RA Mukherjee J.J., Dekker E.E.;
 RT "Purification, properties, and N-terminal amino acid sequence of
 homogeneous Escherichia coli 2-amino-3-ketobutyrate CoA ligase, a
 pyridoxal phosphate-dependent enzyme".
 RL J. Biol. Chem. 262:14441-14447(1987).
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + GLYCINE = COA + 2-AMINO-
 3-OXOBUTANOATE.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 AMINOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL: X06690; CAA29883.1; -
 DR EMBL: U00039; AAB18594.1; -
 DR EMBL: AE000439; AAC76641.1; -
 DR EMBL: AE005589; AAG58764.1; -
 DR EMBL: AP002566; BAB37918.1; -
 DR PIR: S00913; XUECGA.
 DR PIR: A29474; A29474.
 DR ECO2DBASE: G042.2; 6TH EDITION.
 DR EcoGene: EG10512; Kbl.
 DR InterPro: IPR001917; AminoTransf_2.
 DR Pfam: PF00222; aminotran_2; 1.
 DR PROSITE: PS00599; AA.TRANSFER.CLASS.2; 1.
 KW Transferase; Acyltransferase; Pyridoxal phosphate; Complete proteome.
 FT BINDING 244 244
 FT CONFLICT 43 43 H->Q (IN REF. 1).
 FT CONFLICT 171 171 A->R (IN REF. 1).
 FT CONFLICT 183 183 V->L (IN REF. 1).
 SQ SEQUENCE 398 AA; 43117 MW; 76E5DC4AA2F84F5 CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 210 GGLFETL 216
 |||||
 DB 116 GGLFETL 122

RESULT 59
 THIB_CANTR STANDARD; PRT; 402 AA.
 ID THIB_CANTR
 AC Q04677;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ACETYL-COA ACETYLTRANSFERASE IB (EC 2.3.1.9) (PEROXISOMAL
 DE ACETOACETYL-COA THIOLASE) (THIOLASE IB).
 DE PACTB.
 GN Candida tropicalis (Yeast).
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5482;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-16; 208-232 AND 279-289.
 RX MEDLINE=93130927; PubMed=1362382;
 RA Kuritani T., Ueda M., Kanayama N., Kondo J., Teranishi Y., Tanaka A.;
 RT "Peroxisomal acetoacetyl-CoA thiolase of an n-alkane-utilizing yeast,
 Candida tropicalis".
 RL Eur. J. Biochem. 210:999-1005(1992).
 CC -1- CATALYTIC ACTIVITY: 2 ACETYL-COA -> COA + ACETOACETYL-COA.
 CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF POLY BETA-HYDROXYBUTYRATE.
 CC -1- SUBUNIT: MULTIMERIC (BY SIMILARITY).

```

CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
CC -1- SIMILARITY: BELONGS TO THE THIOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D13471; BAA02716.1; -
CC PIR: S28144; S28144.
CC HSSP: P27796; 1PXT.
CC InterPro: IPR002155; Thiolase.
CC Pfam: PF00108; Thiolase_1.
CC PROSITE: PS00098; THIOLEASE_1; 1.
CC PROSITE: PS00099; THIOLEASE_3; 1.
CC PROSITE: PS00737; THIOLEASE_2; 1.
CC PROSITE: PS00342; MICROBODIES_CTER; 1.
CC Transferase: Acyltransferase; Peroxisome.
CC INT MET 0 0
CC ACT_SITE 90 90 SUBSTRATE BINDING (BY SIMILARITY).
CC SITE 382 382 MICROBODY TARGETING SIGNAL (POTENTIAL).
CC SITE 400 402
CC SEQUENCE 402 AA; 41799 MW; 599BAF2D195F6757 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 893 QGSLSL 899
    |||||
Db 19 QGSLSL 25

RESULT 60
Y097_AQUAE STANDARD: PRT; 407 AA.
AC 066504:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_097.
GN AQ_097.
OS Aquifex aeolicus.
CC Bacteria; Aquificales; Aquificaceae; Aquifex.
CX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Girdham R.A., Overbeek R., Sneed M.A., Kellier R.V., Huber R.,
RA Feldman D.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000673; AAC06469.1; -
CC KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 40 62 POTENTIAL.

```

```

FT TRANSMEM 67 89 POTENTIAL.
FT TRANSMEM 118 140 POTENTIAL.
FT TRANSMEM 147 169 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
FT TRANSMEM 253 271 POTENTIAL.
FT TRANSMEM 281 303 POTENTIAL.
FT TRANSMEM 334 356 POTENTIAL.
FT TRANSMEM 361 378 POTENTIAL.
FT TRANSMEM 385 402 POTENTIAL.
SQ SEQUENCE 407 AA; 46104 MW; E3C8738086B343BF CRC64;

Query Match 0.7%; Score 7; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 LTTGGL 427
    |||||
Db 151 LTTGGL 157

RESULT 61
PGK_PYRMO STANDARD: PRT; 410 AA.
ID PGK_PYRMO
AC P50316;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN PGK.
OS Pyrococcus woesei.
CC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
CX NCBI_TaxID=2262;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=DSM 3773;
RX MEDLINE=96061954; PubMed=7588750;
RA Hess D., Krueger K., Knappik A., Palm P., Hensel R.;
RA "Dimeric 3-phosphoglycerate kinases from hyperthermophilic Archaea.
RT Cloning, sequencing and expression of the 3-phosphoglycerate kinase
RT gene of Pyrococcus woesei in Escherichia coli and characterization of
RT the protein. Structural and functional comparison with the 3-
RL phosphoglycerate kinase of Methanothermus fervidus";
RL Eur. J. Biochem 233:227-237(1995).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73527; CA51930.1; -
CC DR HSSP: P36204; 1YPE.
CC DR InterPro: IPR001576; PGK.
CC DR Pfam: PF00162; PGK; 1.
CC DR PRINTS: PR00477; PHGLYCKINASE.
CC DR PROSITE: PS00111; GLYCERATE_KINASE; 1.
CC KW Transferase; Kinase; Glycolysis.
SQ SEQUENCE 410 AA; 46224 MW; D90D4172A06707E0 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 410;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 984 VRKLSQV 990
 DB 134 VRKLSQV 140

RESULT 62

2ABA_PIG STANDARD; PRT; 426 AA.
 AC 029090;
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SERINE/THREONINE PROTEIN PHOSPHATASE 2A, 55 KDA REGULATORY SUBUNIT B,
 DE ALPHA ISOFORM (PP2A, SUBUNIT B, B-ALPHA ISOFORM) (PP2A, SUBUNIT B,
 DE B55-ALPHA ISOFORM) (PP2A, SUBUNIT B, PP55-ALPHA ISOFORM) (PP2A,
 DE SUBUNIT B, R2-ALPHA ISOFORM) (FRAGMENT).
 GN PPP2R2A.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mayer-Jeckel R.E.;
 RL Theiss (1992), Friedrich Miescher Institut / Basel, Switzerland.
 CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
 CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
 CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
 CC COMPARTMENT.
 CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
 CC COMPOSED OF A 36 KDA CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 KDA
 CC CONSTANT REGULATORY SUBUNIT (PP65 OR SUBUNIT A), THAT ASSOCIATES
 CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
 CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
 CC B (THE R2/B/PP55/B55, R3/B'/PP72/PP130/PP59 AND R5/B'/B56
 CC FAMILIES), THE 48 KDA VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
 CC AND CELL SIGNALING MOLECULES.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 234932; CAA84404.1; -
 CC InterPro; IPR000009; PP2A_PP55.
 CC DR InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40.5.
 CC DR PRINTS; PR00600; PP2APR55.
 CC SMART; SM00320; WD40.3.
 CC DR PROSITE; PS01024; PR55_1; 1.
 CC DR PROSITE; PS01025; PR55_2; 1.
 CC KW Multigene family.
 CC FT NON_TER 1
 CC SQUENCE 426 AA; 49613 MW; 3AAD7EB338B03534 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 426;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 717 LMWEASP 723
 DB 141 LMWEASP 147

RESULT 63
 PHOR_ECOLI STANDARD; PRT; 431 AA.
 ID PHOR_ECOLI

AC P08400;
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-AUG-1988 (Rel. 08, Last annotation update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHATE REGULATOR SENSOR PROTEIN PHOR (EC 2.7.3.-).
 GN PHOR OR NMPB OR B0400.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-K12;
 RC MEDLINE=87169739; PubMed=3550103;
 RX Makino K., Shinagawa H., Amenura M., Nakata A.;
 RT "Nucleotide sequence of the phor gene, a regulatory gene for the
 RT phosphate regulation of Escherichia coli.";
 RL J. Mol. Biol. 192:549-556(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1233-1238(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
 RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=90251245; PubMed=2187152;
 RA Yamada M., Makino K., Shinagawa H., Nakata A.;
 RT "Regulation of the phosphate regulation of Escherichia coli: properties
 RT of phor deletion mutants and subcellular localization of phor
 RT protein.";
 RL Mol. Gen. Genet. 220:366-372(1990).
 RN [5]
 RP TOPOLOGY.
 RX MEDLINE=93302503; PubMed=8391104;
 RA Scholten M., Tomassen J.;
 RT "Topology of the phor protein of Escherichia coli and functional
 RT analysis of internal deletion mutants.";
 RL Mol. Microbiol. 8:269-275(1993).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PHOR/PHOB
 CC INVOLVED IN THE PHOSPHATE REGULATION GENES EXPRESSION. PHOR MAY
 CC FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
 CC PHOSPHORYLATES PHOB IN RESPONSE TO ENVIRONMENTAL SIGNALS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSOR TRANSDUCTION HISTIDINE
 CC KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X04704; CAA28409.1; -
 CC EMBL; AE000146; AAC73503.1; -
 CC DR EMBL; U73857; AAB18124.1; -
 CC PIR; A25557; RGECPH.
 CC PIR; S11888; S11888.
 CC Ecogene; EG10733; PHOR.
 CC DR Ecogene; IPR000410; Bctr1_sensor.
 CC InterPro; IPR003594; HATPase_C.
 CC DR


```

DR PRINTS: PR00318; GPROTEINA.
DR SMART: SM00275; G-alpha; 1.
KW GTP-binding; Transducer; Myristate.
FT INIT_MER 0 0 BY SIMILARITY.
FT LIPID 1 1 MYRISTATE (BY SIMILARITY).
FT NP_BIND 118 125 GTP (BY SIMILARITY).
FT NP_BIND 279 283 GTP (BY SIMILARITY).
FT NP_BIND 348 351 GTP (BY SIMILARITY).
FT CONFLICT 22 22 T -> S (IN REF. 2).
FT CONFLICT 372 375 MRRN -> GAD (IN REF. 2).
SQ SEQUENCE 432 AA; 47848 MW; FA0E9BBE5071A8D CRC64;

Query Match 0.7%; Score 7; DB 1; Length 432;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 DSAQALI 309
    |||||
Db 157 DSAQALI 163

RESULT 66
CXAB_SHEEP STANDARD; PRT; 439 AA.
ID AC P55917; Q9MYL3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GAP JUNCTION ALPHA-8 PROTEIN (CONNEXIN 49) (CX49) (LENS FIBER PROTEIN
DE MP70) (MP64) (MP38).
GN GJ48
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP MEDLINE=96254875; PubMed=8654111;
RX Yang D.-I., Louis C.F.;
RT "Molecular cloning of sheep connexin49 and its identity with MP70.";
RL Curr. Eye Res. 15:307-314(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Yang D.-I., Louis C.F.;
RA Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-20.
RC TISSUE=Lens;
RX MEDLINE=88143159; PubMed=2830542;
RA Kistler J., Christie D., Bullivant S.;
RT "Homologies between gap junction proteins in lens, heart and liver.";
RL Nature 331:721-723(1988).
CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EYE LENS.
CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. ALPHA-TYPE (GROUP II)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S82364; AAB37689.1; -
CC DR EMBL; AF177913; AAF01367.1; -

```

```

DR InterPro; IPR000500; Connexin.
DR Pfam; PF00029; connexin; 1.
DR PRINTS: PR00206; CONNEXIN.
DR PRINTS: PR01137; CONNEXINAB.
DR SMART; SM00377; CNX; 1.
DR PROSITE; PS00407; CONNEXINS_1; 1.
DR PROSITE; PS00408; CONNEXINS_2; 1.
DE Gap junction; Transmembrane.
FT INIT_MER 0 0
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 43 POTENTIAL.
FT DOMAIN 44 75 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 76 96 POTENTIAL.
FT DOMAIN 97 160 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 161 181 POTENTIAL.
FT DOMAIN 182 209 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 210 230 POTENTIAL.
FT DOMAIN 231 439 CYTOPLASMIC (POTENTIAL).
FT VARIANT 8 9 N -> R.
FT VARIANT 9 9 I -> L.
FT VARIANT 13 13 V -> A.
FT VARIANT 13 13 E -> N (IN REF. 3).
FT CONFLICT 12 12 N -> Q (IN REF. 3).
FT CONFLICT 14 14 E -> D (IN REF. 2).
FT CONFLICT 41 41
SQ SEQUENCE 439 AA; 49028 MW; 3C941E2BD0332A7E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 718 MVEASPL 724
    |||||
Db 292 MVEASPL 298

RESULT 67
2ABA_HUMAN STANDARD; PRT; 447 AA.
ID AC Q00007; P50409;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SERINE/THREONINE PROTEIN PHOSPHATASE 2A, 55 KDA REGULATORY SUBUNIT B,
DE ALPHA ISOFORM (PP2A, SUBUNIT B, B-ALPHA ISOFORM) (PP2A, SUBUNIT B,
DE B55-ALPHA ISOFORM) (PP2A, SUBUNIT B, PR55-ALPHA ISOFORM) (PP2A,
DE SUBUNIT B, R2-ALPHA ISOFORM).
GN PPP2R2A.
OS Homo sapiens (Human), and Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606, 9986;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Human; TISSUE=Lung fibroblast;
RX MEDLINE=91198016; PubMed=1849734;
RA Mayer R.E., Hendrix P., Cron P., Mathies R., Stone S.R., Goris J.,
RA Merlieve W., Hofsteenge J., Hemmings B.A.;
RT "Structure of the 55-kDa regulatory subunit of protein phosphatase
RT 2A: evidence for a neuronal-specific isoform.";
RL Biochemistry 30:3589-3597(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Rabbit; STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;
RA Depaulis-Roach A.A.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC COMPARTMENT.
CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC COMPOSED OF A 36 KDA CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 KDA
CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES

```

```

CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
CC FAMILIES), THE 48 KDA VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC AND CELL SIGNALING MOLECULES.
CC -1- TISSUE SPECIFICITY: IN ALL TISSUES EXAMINED.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M6A929; AAA36490.1; -
DR EMBL: U09356; AAA18497.1; -
DR PIR: A38351; A38351.
DR MIM: 604941; -
DR InterPro: IPR000009; PP2A_PR55.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR PRINTS: PR00600; PP2APR55.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS01024; PR55_1; 1.
DR PROSITE: PS01025; PR55_2; 1.
DR MultiGene family.
FT MOD_RES 21 21 BLOCKED.
SQ SEQUENCE 447 AA; 51692 MW; PAD407FF7ADA4ED6 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 717 LMVEASP 723
Db 162 LMVEASP 168

RESULT 68
2ABD_RAT STANDARD; PRT; 447 AA.
ID 2ABD_RAT P36876; O35512;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SERINE/THREONINE PROTEIN PHOSPHATASE 2A, 55 KDA REGULATORY SUBUNIT B,
DE ALPHA ISOFORM (PP2A, SUBUNIT B, B-ALPHA ISOFORM) (PP2A, SUBUNIT B,
DE B55-ALPHA ISOFORM) (PP2A, SUBUNIT B, PR55-ALPHA ISOFORM) (PP2A,
DE SUBUNIT B, R2-ALPHA ISOFORM) (PP2A, SUBUNIT B, BRA ISOFORM).
GN PPR2R2A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114192; PubMed=1370560;
RA Pallas D.C., Weller W., Jaspers S., Miller T.B. Jr., Lane W.S.,
RA Roberts T.M.;
RT "The third subunit of protein phosphatase 2A (PP2A), a 55-kilodalton
RT protein which is apparently substituted for by T antigens in
RT complexes with the 36- and 63-kilodalton PP2A subunits, bears little
RT resemblance to T antigens."
RL J. Virol. 66:886-893(1992).
RN [2]
RP SEQUENCE OF 80-272 FROM N.A.
RC STRAIN=FISCHER 344;
RX MEDLINE=93279382; PubMed=8389301;
RA Hatano Y., Shima H., Haneji T., Miura A.B., Sugimura T., Nagao M.;

```

```

RT "Expression of PP2A B regulatory subunit beta isoform in rat
RT testis."
RL FEBS Lett. 324:71-75(1993).
CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC COMPARTMENT.
CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC COMPOSED OF A 36 KDA CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 KDA
CC CONSTANT REGULATORY SUBUNIT (PR55 OR SUBUNIT A), THAT ASSOCIATES
CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
CC FAMILIES), THE 48 KDA VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC AND CELL SIGNALING MOLECULES.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M83298; AAA41910.1; -
DR EMBL: M83297; AAA41909.1; -
DR EMBL: D14419; BAA21904.1; -
DR PIR: A41805; A41805.
DR InterPro: IPR000009; PP2A_PR55.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR PRINTS: PR00600; PP2APR55.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS01024; PR55_1; 1.
DR PROSITE: PS01025; PR55_2; 1.
DR MultiGene family.
KW VARIANT.
FT VARIANT 60 60 E -> ESFKVAALREASNSMQ.
FT CONFLICT 105 105 K -> R (IN REF. 1; AAA41909).
FT CONFLICT 105 105 K -> R (IN REF. 2).
FT CONFLICT 213 213 N -> S (IN REF. 2).
FT CONFLICT 222 222 M -> V (IN REF. 2).
SQ SEQUENCE 447 AA; 51678 MW; 180AC837D9DA4ECE CRC64;

Query Match 0.7%; Score 7; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 717 LMVEASP 723
Db 162 LMVEASP 168

RESULT 69
2ABD_RAT STANDARD; PRT; 453 AA.
ID 2ABD_RAT P56932;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SERINE/THREONINE PROTEIN PHOSPHATASE 2A, 55 KDA REGULATORY SUBUNIT B,
DE DELTA ISOFORM (PP2A, SUBUNIT B, B-DELTA ISOFORM) (PP2A, SUBUNIT B,
DE B55-DELTA ISOFORM) (PP2A, SUBUNIT B, PR55-DELTA ISOFORM) (PP2A,
DE SUBUNIT B, R2-DELTA ISOFORM).
GN PPR2R2D.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Brain;
RX MEDLINE=20026081; PubMed=10556517;
RA Strack S., Chang D., Zaucha J.A., Colbran R.J., Wadzinski B.E.;
RT "Cloning and characterization of B delta, a novel regulatory subunit
  of protein phosphatase 2A."
RL FEBS Lett. 460:462-466(1999).
CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
  SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
  LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
  COMPARTMENT.
CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
  COMPOSED OF A 36 KDA CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 KDA
  CONSTRAINT REGULATORY SUBUNIT (PRE5 OR SUBUNIT A), THAT ASSOCIATES
  WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
  WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
  B (THE R2/B/PR5/955, R3/B'/PR2/PR130/PR59 AND R5/B'/B56
  FAMILIES), THE 48 KDA VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
  AND CELL SIGNALING MOLECULES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGH LEVELS IN BRAIN,
  HEART, PLACENTA, SKELETAL MUSCLE, TESTIS, THYMUS AND SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
  FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF180350; AAF08536.1; -
DR InterPro: IPR000009; PP2A_PR55.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR SMART: SM00320; WD40; 2.
KM Multigene family.
FT DOMAIN 3 8 POLY-GLY.
SQ SEQUENCE 453 AA; 51982 MW; 733E80A93A5BC2BB CRC64;

Query Match 0.7%; Score 7; DB 1; Length 453;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 717 LMEVESP 723
Db 168 LMEVESP 174

RESULT 70
ID LEFT_KLUFA STANDARD; PRT; 469 AA.
AC P53938;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE LEFT1 PROTEIN.
GN LEFT1.
OS Kluyveromyces lactic (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RA Mesolowski-Louvet M., Tanguy-Rougeau C., Fukuhau H.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YNL240C.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70373; CAA49833.1; -
SQ SEQUENCE 469 AA; 52072 MW; 555DAAE7DEFB9FB7 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 LVRLSQ 989
Db 341 LVRLSQ 347

RESULT 71
ID Y213_MYCGE STANDARD; PRT; 471 AA.
AC P47455;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG213.
GN MG213.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
  Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
  Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
  Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
  Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
  Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Luetter T.S.,
  Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
  RA "The minimal gene complement of Mycoplasma genitalium."
  RT Science 270:397-403(1995).
RL Science 270:397-403(1995).
CC -1- SIMILARITY: SOME, TO B.SUBTILIS YPUG.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39701; AAC71432.1; -
DR TIGR: MG213; -
DR InterPro: IPR003768; DUF173.
DR Pfam: PF02616; DUF173; 1.
KM Hypothetical protein; Complete proteome.
FT DOMAIN 422 427 POLY-GLN.
SQ SEQUENCE 471 AA; 54829 MW; 9531DEC1D790FE2 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 454 RLSSLT 460
Db 316 RLSSLT 322

```

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration


```

RESULT 72
LIPL_MOUSE STANDARD; PRT; 474 AA.
ID LIPL_MOUSE
AC P1152;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LIPOPROTEIN LIPASE PRECURSOR (EC 3.1.1.34) (LPL).
GN LPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112227; PubMed=1765386;
RA Zechner R., Newman T.C., Steiner E., Breslow J.L.;
RT "The structure of the mouse lipoprotein lipase gene: a Bl repetitive
RT element is inserted into the 3' untranslated region of the mRNA.";
RL Genomics 11:62-76(1991).
RN [2]
RP SEQUENCE OF 1-183 FROM N.A.
RX MEDLINE=93054974; PubMed=1339460;
RA Gimble J.M., Hua X., Youkhana K., Bass H.W., Medina K.,
RA Sullivan M., Greendberger J.S., Wang C.S.;
RT "Adipogenesis in a myeloid supporting bone marrow stromal cell line.";
RL J. Cell. Biochem. 50:73-82(1992).
RN [3]
RP SEQUENCE OF 9-474 FROM N.A.
RX MEDLINE=87250454; PubMed=3597382;
RA Kirchgesner T.G., Svenson K.L., Lusis A.J., Schotz M.C.;
RT "The sequence of cDNA encoding lipoprotein lipase. A member of a
RT lipase gene family.";
RL J. Biol. Chem. 262:8463-8466(1987).
CC -1- FUNCTION: THE PRIMARY FUNCTION OF THIS LIPASE IS THE HYDROLYSIS
CC OF TRIGLYCERIDES OF CIRCULATING CHYLOMICRONS AND VERY LOW DENSITY
CC LIPOPROTEINS (VLDL). THE ENZYME FUNCTIONS IN THE PRESENCE OF
CC APOLIPOPROTEIN C-2 ON THE LUMINAL SURFACE OF VASCULAR ENDOTHELIUM.
CC -1- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC A FATTY ACID ANION.
CC -1- SUBUNIT: HOMODIMER, INTERACT WITH APOLIPOPROTEIN C-2.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AAA39441.1; -
DR EMBL: M60838; AAA39441.1; JOINED.
DR EMBL: M60839; AAA39441.1; JOINED.
DR EMBL: M60840; AAA39441.1; JOINED.
DR EMBL: M60842; AAA39441.1; JOINED.
DR EMBL: M60843; AAA39441.1; JOINED.
DR EMBL: M60844; AAA39441.1; JOINED.
DR EMBL: M60845; AAA39441.1; JOINED.
DR EMBL: M60846; AAA39441.1; JOINED.
DR EMBL: M65258; AAA39442.1; -
DR EMBL: J03302; AAA39440.1; -
DR PIR: A29300; A29300.
DR HSSP: P06857; 1RP1.
DR MGD: MGI:96820; LPL.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR InterPro: IPR001024; LH2.
DR InterPro: IPR000734; Lipase.
DR Pfam: PF00151; lipase_1.
DR Pfam: PF01477; PLAT_1.
DR PRINTS: PR00821; TAGLIPASE.

```

```

DR PRINTS: PR00822; LIPO_LIPASE.
DR SMART: SM00308; LH2_1.
DR PROSITE: PS00120; LIPASE_SER.1.
KW Hydrolase; Plasma; Glycoprotein; Lipid degradation; Chylomicron;
KW VLDL; Heparin-binding; GPI-anchor; Signal.
FT CHAIN 1 474 LIPOPROTEIN LIPASE.
FT SIGNAL 27 POTENTIAL.
FT ACT_SITE 159 159 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 183 183 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 268 268 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 319 331 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 54 67 BY SIMILARITY.
FT DISULFID 243 266 BY SIMILARITY.
FT DISULFID 291 310 BY SIMILARITY.
FT DISULFID 302 305 BY SIMILARITY.
FT DISULFID 445 465 BY SIMILARITY.
FT CARBOHYD 70 70 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 147 147 N->K (IN REF. 1).
SQ SEQUENCE 474 AA; 53126 MW; E7AF969E716F1CDD CRC64;

```

```

Query Match 0.7%; Score 7; DB 1; Length 474;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 698 AGVAGSL 704
Db 165 AGVAGSL 171

```

```

RESULT 73
LIPL_RAT STANDARD; PRT; 474 AA.
ID LIPL_RAT
AC Q06000;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LIPOPROTEIN LIPASE PRECURSOR (EC 3.1.1.34) (LPL).
GN LPL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
RX MEDLINE=93077037; PubMed=1339374;
RA Brault D., Noe L., Etienne J., Hamelin J., Raisonnier A.,
RA Souli A., Chuat J.-C., Dugail I., Quignard-Boulangé A., Lavan M.,
RA Gallibert F.;
RT "Sequence of rat lipoprotein lipase-encoding cDNA.";
RL Gene 121:237-246(1992).
CC -1- FUNCTION: THE PRIMARY FUNCTION OF THIS LIPASE IS THE HYDROLYSIS
CC OF TRIGLYCERIDES OF CIRCULATING CHYLOMICRONS AND VERY LOW DENSITY
CC LIPOPROTEINS (VLDL). THE ENZYME FUNCTIONS IN THE PRESENCE OF
CC APOLIPOPROTEIN C-2 ON THE LUMINAL SURFACE OF VASCULAR ENDOTHELIUM.
CC -1- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC A FATTY ACID ANION.
CC -1- SUBUNIT: HOMODIMER, INTERACT WITH APOLIPOPROTEIN C-2.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L03294; AAA1534.1; -

```


DR PIR: JH0790; JH0790.
 DR HSSP: P06857; 1RP1.
 DR InterPro: IPR000379; Est_Lip_thioest_actsite.
 DR InterPro: IPR001024; LH2.
 DR InterPro: IPR00734; Lipase.
 DR Pfam: PF00151; Lipase; 1.
 DR Pfam: PF01477; PLAT; 1.
 DR PRINTS: PR00821; TAGLIPASE.
 DR PRINTS: PR00822; LIPOLIPIASE.
 DR SMART: SM00308; LH2; 1.
 DR PROSITE: PS00120; LIPASE_SER; 1.
 DR Hydrolase; Plasma; glycoprotein; lipid degradation; Chylomicron;
 KW VLDL; Heparin-binding; GPI-anchor; signal.
 FT SIGNAL 1 27
 FT CHAIN 28 474
 FT ACT_SITE 159 159
 FT ACT_SITE 183 183
 FT ACT_SITE 268 268
 FT DOMAIN 319 331
 FT DISULFID 54 67
 FT DISULFID 243 266
 FT DISULFID 291 310
 FT DISULFID 302 305
 FT DISULFID 445 465
 FT CARBOHYD 70 70
 FT CARBOHYD 386 386
 FT SEQUENCE 474 AA; 53082 MW; F4P6F4BCA4F1626 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 474;
 Best local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 698 AGVAGSL 704
 |||||
 Db 165 AGVAGSL 171

RESULT 74
 ODB2_BOVIN STANDARD; PRT; 482 AA.

AC P1181;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE LIPOAMIDE ACYLTRANSFERASE COMPONENT OF BRANCHED-CHAIN ALPHA-KETO ACID
 DE DEHYDROGENASE COMPLEX, MITOCHONDRIAL PRECURSOR (EC 2.3.1.-) (E2)
 DE (DIHYDROLIPOAMIDE BRANCHED CHAIN TRANSACYLASE) (BCKAD E2 SUBUNIT).
 GN DBT.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89008232; PubMed=3049570;
 RA Griffin T.A., Lau K.S., Chuang D.T.;
 RT "Characterization and conservation of the inner E2 core domain
 RT structure of branched-chain alpha-keto acid dehydrogenase complex
 RT from bovine liver. Construction of a cDNA encoding the entire
 RT transacylase (E2b) precursor.";
 RT J. Biol. Chem. 263:14008-14014(1988).
 RP [2]
 RP SEQUENCE OF 1-227 FROM N.A.
 RX MEDLINE=88241022; PubMed=2837277;
 RA Lau K.S., Griffin T.A., Hu C.-W.C., Chuang D.T.;
 RT "Conservation of primary structure in the lipoyl-bearing and
 RT dihydrolipoyl dehydrogenase binding domains of mammalian
 RT branched-chain alpha-keto acid dehydrogenase complex: molecular
 RT cloning of human and bovine transacylase (E2) cDNAs.";
 RT Biochemistry 27:1972-1981(1988).
 CC -!- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX

CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
 CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:
 CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
 CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).
 CC COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
 CC -!- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
 CC SYMMETRY.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -!- MISCELLANEOUS: THE CATALYTIC FUNCTION OF THIS ENZYME IS TO ACCEPT,
 CC AND TO TRANSFER TO COENZYME A, ACYL GROUPS THAT ARE GENERATED BY
 CC THE BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE COMPONENT.
 CC -!- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M21572; AAA30597.1; -
 DR EMBL: M19475; AAA30596.1; ALT_SEQ.
 DR PIR: A30801; XUBOLA.
 DR PIR: B28707; B28707.
 DR HSSP: P11961; 2PDD.
 DR InterPro: IPR001078; 2oxoacid_dh.
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR InterPro: IPR003016; Lipoyl.
 DR Pfam: PF00198; 2-oxoacid_dh; 1.
 DR Pfam: PF00364; biotin_lipoyl; 1.
 DR ProDom: PD001115; 2oxoacid_dh; 1.
 DR PROSITE: PS00189; Lipoyl; 1.
 DR TRANSFERASE: Acyltransferase; Mitochondrion; Transf. peptide; Lipoyl.
 FT TRANSIT 1 61
 FT CHAIN 62 482
 FT BINDING 105 105
 FT ACT_SITE 452 452
 FT ACT_SITE 456 456
 FT CONFLICT 32 32
 FT CONFLICT 173 173
 FT SEQUENCE 482 AA; 53410 MW; 70F06E6E2814B5C6 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 482;
 Best local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 801 LSDIGEG 807
 |||||
 Db 69 LSDIGEG 75

RESULT 75
 ODB2_HUMAN STANDARD; PRT; 482 AA.

AC P1182;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LIPOAMIDE ACYLTRANSFERASE COMPONENT OF BRANCHED-CHAIN ALPHA-KETO ACID
 DE DEHYDROGENASE COMPLEX, MITOCHONDRIAL PRECURSOR (EC 2.3.1.-) (E2)
 DE (DIHYDROLIPOAMIDE BRANCHED CHAIN TRANSACYLASE) (BCKAD E2 SUBUNIT).
 GN DBT OR BCARE2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=93041936; PubMed=1420314;
 RA Lau K.S., Chuang J.L., Herring W.J., Danner D.J., Cox R.P.,
 RT Chuang D.T.;
 RT "The complete cDNA sequence for dihydrolipoyl transacylase (E2) of
 human branched-chain alpha-keto acid dehydrogenase complex.";
 RL Biochim. Biophys. Acta 1132:319-321(1992).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=88198156; PubMed=3245861;
 RA Hummel K.B., Litwer S., Bradford A.P., Aitken A., Danner D.J.,
 RT Yeaman S.J.;
 RT "Nucleotide sequence of a cDNA for branched chain acyltransferase
 with analysis of the deduced protein structure.";
 RL J. Biol. Chem. 263:6165-6168(1988).
 [3]
 RN REVISIONS.
 RX MEDLINE=89214230; PubMed=2708389;
 RA Danner D.J., Litwer S., Herring W.J., Pruckler J.;
 RT "Construction and nucleotide sequence of a cDNA encoding the
 full-length preprotein for human branched chain acyltransferase.";
 RL J. Biol. Chem. 264:7742-7746(1989).
 [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=89302075; PubMed=2742576;
 RA Nobukuni Y., Mitsubuchi H., Endo F., Matsuda I.;
 RT "Complete primary structure of the transacylase (E2b) subunit of the
 human branched chain alpha-keto acid dehydrogenase complex.";
 RL Biochem. Biophys. Res. Commun. 161:1035-1041(1989).
 [5]
 RN SEQUENCE OF 1-313 FROM N.A.
 RX MEDLINE=88241022; PubMed=2837277;
 RA Lau K.S., Griffin T.A., Hu C.-W.C., Chuang D.T.;
 RT "Conservation of primary structure in the lipoyl-bearing and
 dihydrolipoyl dehydrogenase binding domains of mammalian
 branched-chain alpha-keto acid dehydrogenase complex: molecular
 cloning of human and bovine transacylase (E2) cDNAs.";
 RL Biochemistry 27:1972-1981(1988).
 [6]
 RN VARIANT MSUD CYS-276.
 RX MEDLINE=91128420; PubMed=1847055;
 RA Fisher C.W., Lau K.S., Fisher C.R., Wynn R.M., Cox R.P., Chuang D.T.;
 RT "A 17-bp insertion and a Phe215->Cys missense mutation in the
 dihydrolipoyl transacylase (E2) mRNA from a thiamine-responsive maple
 syrup urine disease patient WC-34.";
 RL Biochem. Biophys. Res. Commun. 174:804-809(1991).
 CC -1- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
 CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
 AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:
 CC BRANCHED-CHAIN ALPHA-KETO DECARBOXYLASE (E1), LIPAMIDE
 CC ACYLTRANSFERASE (E2) AND LIPAMIDE DEHYDROGENASE (E3).
 CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
 CC COFACTOR.
 CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
 CC SYMMETRY.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1- DISEASE: DEFECTS IN THE SUBUNITS OF THE BECKO COMPLEX ARE THE
 CC CAUSE OF MAPLE SYRUP URINE DISEASE (MSUD); AN AUTOSOMAL RECESSIVE
 CC DISORDER CHARACTERIZED BY MENTAL AND PHYSICAL RETARDATION, FEEDING
 CC PROBLEMS, AND A MAPLE SYRUP ODOR TO THE URINE.
 CC -1- MISCELLANEOUS: THE CATALYTIC FUNCTION OF THIS ENZYME IS TO ACCEPT,
 CC AND TO TRANSFER TO COENZYME A, ACYL GROUPS THAT ARE GENERATED BY
 CC THE BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE COMPONENT.
 CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN:
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL: X66785; CAA47285.1; -;
 DR EMBL: J03208; AAA35589.1; ALT_INIT.
 DR EMBL: M27093; AAA64512.1; ALT_INIT.
 DR EMBL: M19301; AAA59200.1; ALT_SEO.
 DR PIR: A28655; A28655.
 DR PIR: A28707; A28707.
 DR PIR: A32422; A32422.
 DR PIR: A33362; A33362.
 DR PIR: S27194; S27194.
 DR PIR: S22865; S22865.
 DR HSSP: P11961; 2PDD.
 DR MIM: 248610; -;
 DR InterPro: IPR001078; 2oxoacid_dh.
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR InterPro: IPR003016; Lipoyl.
 DR Pfam: PF00198; 2-oxoacid_dh; 1.
 DR Pfam: PF00364; biotin_lipoyl; 1.
 DR ProDom: PD001115; 2oxoacid_dh; 1.
 DR ProSite: PS00189; Lipoyl; 1.
 KW Transferrase: Acyltransferase; Mitochondrion; Transit peptide; Lipoyl;
 KW Disease mutation: Maple syrup urine disease.
 KW TRANSIT
 FT CHAIN 59 482 MITOCHONDRION (POTENTIAL).
 FT
 FT BINDING 105 105 LIPAMIDE ACYLTRANSFERASE COMPONENT OF
 FT
 FT ACT_SITE 452 452 BRANCHED-CHAIN ALPHA-KETO ACID
 FT
 FT ACT_SITE 456 456 DEHYDROGENASE COMPLEX.
 FT
 FT VARIANT 276 276 LIPOYL (POTENTIAL).
 FT
 FT CONFLICT 321 321 F->C (IN MSUD).
 FT
 FT CONFLICT 354 354 Q->P (IN REF. 4).
 FT
 FT SEQUENCE 482 AA; 53487 MW; ATCA728C8F33D126 CRC64; L->V (IN REF. 4).
 SQ

Query Match 0.7%; Score 7; DB 1; Length 482;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Job time: 210 sec

Oy 801 LSDIGEG 807
 Db 69 LSDIGEG 75

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 10:59:10 ; Search time 21 Seconds

(without alignments)
3714.416 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 1024

Sequence: 1 MNFKDNRSLIORMGMTVI.....MQFDDDLSTVTGAFLVTA 1024

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.9	414	2 A36443	seminal vesicle se
2	9	0.9	508	2 S11863	aldenhyde dehydroge
3	9	0.9	1021	2 E64576	hypothetical prote
4	9	0.9	1120	2 B86479	hypothetical prote
5	8	0.8	111	2 T40946	60s ribosomal prot
6	8	0.8	112	2 T38228	60s ribosomal prot
7	8	0.8	138	2 S66004	transcription regu
8	8	0.8	221	2 T07079	leucine-rich repea
9	8	0.8	225	1 ADPSGP	2-dehydro-3-deoxyp
10	8	0.8	280	2 C86317	protein T10022.23
11	8	0.8	350	2 T28975	hypothetical prote
12	8	0.8	356	2 E81269	hypothetical prote
13	8	0.8	368	2 F81816	phosphoserine tran
14	8	0.8	368	2 H81059	phosphoserine tran
15	8	0.8	403	1 GRECY	lysine-specific
16	8	0.8	403	2 D85804	lysine-specific
17	8	0.8	455	2 B70664	hypothetical prote
18	8	0.8	466	2 JH0197	muscarinic acetylch
19	8	0.8	466	2 S10126	muscarinic acetylch
20	8	0.8	466	2 S10856	muscarinic acetylch
21	8	0.8	466	2 A27386	glucuronosyltransf
22	8	0.8	535	2 I57961	hypothetical prote
23	8	0.8	680	2 E85431	hypothetical prote
24	8	0.8	806	2 T18840	hypothetical prote
25	8	0.8	1112	2 T10504	disease resistance
26	8	0.8	1200	2 T43148	probable protein-t
27	7	0.7	72	2 E70763	hypothetical prote
28	7	0.7	83	2 E65027	hypothetical prote
29	7	0.7	96	2 S35715	glycine-rich prote

30	7	0.7	98	2 H71630	hypothetical prote
31	7	0.7	103	2 D72710	hypothetical prote
32	7	0.7	104	2 S04103	hemolysin A - Morg
33	7	0.7	107	4 S57388	hypothetical prote
34	7	0.7	111	2 B82830	hypothetical prote
35	7	0.7	118	2 S38608	ribosomal protein
36	7	0.7	133	2 A29774	T-cell receptor al
37	7	0.7	138	2 D72539	hypothetical prote
38	7	0.7	144	2 S04069	glycine-rich prote
39	7	0.7	146	2 S35716	glycine-rich prote
40	7	0.7	146	2 T48221	hypothetical prote
41	7	0.7	147	2 H83076	probable peptidyl-T
42	7	0.7	157	2 S25499	cell receptor al
43	7	0.7	157	2 B82867	hypothetical prote
44	7	0.7	169	1 MORL2	myosin L2 (DTRB) r
45	7	0.7	170	1 MORBLD	myosin L2 (DTRB) r
46	7	0.7	170	2 S12855	myosin L2 regulato
47	7	0.7	171	2 JH0783	diamine N-acetyltr
48	7	0.7	171	2 S25724	hypothetical prote
49	7	0.7	178	2 E69411	conserved hypothet
50	7	0.7	181	2 S20528	hypothetical prote
51	7	0.7	186	2 B64398	hypothetical prote
52	7	0.7	198	2 S48994	hypothetical prote
53	7	0.7	210	2 F83751	ABC transporter (A
54	7	0.7	221	2 G81335	probable ABC trans
55	7	0.7	222	2 F71309	probable phosphogl
56	7	0.7	226	2 B69670	choline ABC transp
57	7	0.7	226	2 T30478	hypothetical prote
58	7	0.7	233	2 C72105	conserved hypothet
59	7	0.7	233	2 D86519	disulfide bond cha
60	7	0.7	235	2 G75494	AziC family protei
61	7	0.7	237	2 T79268	trac protein - Esc
62	7	0.7	240	2 A64553	probable holooytic
63	7	0.7	242	2 E70959	hypothetical prote
64	7	0.7	256	2 G84061	ferrichrome ABC tr
65	7	0.7	259	2 T18151	hypothetical prote
66	7	0.7	264	2 T36431	probable ABC-type
67	7	0.7	264	2 A69102	trehalose-6-phosph
68	7	0.7	264	2 E72642	hypothetical prote
69	7	0.7	265	2 S46534	ubiquinol--cytochr
70	7	0.7	266	2 T29609	hypothetical prote
71	7	0.7	266	2 A72401	ABC transporter, A
72	7	0.7	268	2 A55511	2-hydroxypenta-2,4
73	7	0.7	271	2 S72382	hypothetical prote
74	7	0.7	273	2 A41607	ubiquinol--cytochr
75	7	0.7	276	2 E65742	ABC transporter (A
76	7	0.7	283	2 F82779	hypothetical prote
77	7	0.7	284	2 T09840	amino acid transpo
78	7	0.7	284	2 G82319	DnaJ-related prote
79	7	0.7	288	2 A31326	T-cell receptor de
80	7	0.7	288	2 F86659	ABC transporter AT
81	7	0.7	301	2 A70335	hypothetical prote
82	7	0.7	305	2 C71853	methionyl--trna for
83	7	0.7	306	2 G81374	ornithine carbamoy
84	7	0.7	306	2 T05605	hypothetical prote
85	7	0.7	308	2 A75186	quinolinate synthe
86	7	0.7	308	2 H75049	hypothetical prote
87	7	0.7	308	2 A36918	hypothetical prote
88	7	0.7	308	2 S71746	ATP-binding cassat
89	7	0.7	309	2 S18999	cwi1 protein - fis
90	7	0.7	322	2 T22620	hypothetical prote
91	7	0.7	322	2 S18999	asparaginase (EC 3
92	7	0.7	326	2 E84635	hypothetical prote
93	7	0.7	332	2 T05094	peroxidase homolog
94	7	0.7	332	2 S49738	probable membrane
95	7	0.7	335	2 S75621	hypothetical prote
96	7	0.7	335	2 E71140	hypothetical prote
97	7	0.7	341	2 F82301	peptide ABC transp
98	7	0.7	345	2 D84012	N-acetylglutamate
99	7	0.7	347	1 S51120	alcohol dehydrogen
100	7	0.7	357	2 B71347	probable catabolit
			2	2 A85675	hypothetical prote

ALIGNMENTS

RESULT 1

A36443

seminal vesicle secretory protein II precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 05-Nov-1999

C:Accession: A36443

R:Harris, S.E.; Harris, M.A.; Johnson, C.M.; Bean, M.F.; Dodd, J.G.; Matusik, R.J.; Carr

J. Biol. Chem. 265, 9896-9903, 1990

A:Title: Structural characterization of the rat seminal vesicle secretion II protein and

A:Reference number: A36443; MUID:90277684

A:Accession: A36443

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-414 <HAR>

A:Cross-references: GB:J05443; NID:g207114; PIDN:AAA42192.1; PID:g207115

Query Match

Best Local Similarity 100.0%; Score 9; DB 2; Length 414;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1006 QPDDDLSTV 1014

Db 335 QPDDDLSTV 343

RESULT 2

S11863

aldehyde dehydrogenase homolog - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jan-2000

C:Accession: S11863

R:Guerrero, F.D.; Jones, J.T.; Mullet, J.E.

Plant Mol. Biol. 15, 11-26, 1990

A:Title: Turgor-responsive gene transcription and RNA levels increase rapidly when pea s

A:Reference number: S11861; MUID:91355842

A:Accession: S11863

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-508 <GUE>

A:Cross-references: EMBL:X54359; NID:920680; PIDN:CAA38243.1; PID:g20681

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:57-320/Domain: aldehyde dehydrogenase homology <ALD>

Query Match

Best Local Similarity 100.0%; Score 9; DB 2; Length 508;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 876 NVLEQRLAL 884

Db 315 NVLEQRLAL 323

RESULT 3

E64576

hypothetical protein HP0453 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 04-Mar-2000

C:Accession: E64576

R:Tomb, J.F.; White, O.; Kellavag, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Khaliq, H.G.; Glodek, A.; McKenney, L.

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467

A:Accession: E64576

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1021 <TOM>

A:Cross-references: GB:AE000560; GB:AE000511; NID:92313554; PIDN:AD07519.1; PID:g231

C:Superfamily: Helicobacter pylori hypothetical protein HP0453

Query Match

Best Local Similarity 100.0%; Score 9; DB 2; Length 1021;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 735 VTNKLTLSTI 743

Db 879 VTNKLTLSTI 887

RESULT 4

B86479

hypothetical protein AAF79881.1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: B86479

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: B86479

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1120 <SNO>

A:Cross-references: GB:AE005172; NID:98778966; PIDN:AAF79881.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match

Best Local Similarity 100.0%; Score 9; DB 2; Length 1120;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 SLGNLKNLT 766

Db 145 SLGNLKNLT 153

RESULT 5

T40946

60s ribosomal protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C:Accession: T40946

R:Lucas, M.; Galliard, C.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, January 1999

A:Reference number: Z21959

A:Accession: T40946

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-111 <LUC>

A:Cross-references: EMBL:AL035359; PIDN:CAA22866.1; GSPDB:GN00068; SPDB:SPCC1322.15

A:Experimental source: strain 972h-; cosmid c1322

C:Genetics:

A:Gene: SPDB:SPCC1322.15

A:Map position: 3

C:Superfamily: rat ribosomal protein L34

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 111;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 831 CLSANAVK 838
 |||||||
 Db 79 CLSANAVK 86

RESULT 6

T38228
 60S ribosomal protein L34 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T38228
 R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1998
 A:Reference number: 221780
 A:Accession: T38228
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-112 <MUR>
 A:Cross-references: EMBL:AL021813; PIDN:CA16982.1; GSPDB:GN00066; SPDB:SPAC23A1.08C
 A:Experimental source: strain 972h; cosmid c23A1
 C:Genetics:
 A:Gene: RPL34; SPDB:SPAC23A1.08C
 A:Map position: 1
 C:Superfamily: rat ribosomal protein L34

Query Match 0.8%; Score 8; DB 2; Length 112;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 831 CLSANAVK 838
 |||||||
 Db 79 CLSANAVK 86

RESULT 7

S66004
 transcription regulator homolog yyan - Bacillus subtilis
 N:Alternate names: probable DNA-binding protein yyan
 C:Species: Bacillus subtilis
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
 C:Accession: S66004; E70085
 R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
 DNA Res. 1, 1-14, 1994
 A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom
 A:Reference number: S65967; MUID:96051385
 A:Accession: S66004
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-138 <OGA>
 A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BA05210.1; PID:g467364
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
 R:Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berlet
 C.; Bron, S.; Bouilliet, S.; Birsich, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 350, 249-256, 1997
 A:Authors: Foulter, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinols,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y, M.; Oqawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivalta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 akuch, M.; Tanakashi, A.; Tanaka, T.; Tepsera, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipaty, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E70085
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-138 <KUN>

A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB16117.1; PID:g26366
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yyan
 C:Superfamily: transcription repressor glnR

Query Match 0.8%; Score 8; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 68 LFLKSLE 75
 |||||||
 Db 51 LFLKSLE 58

RESULT 8

T07079
 leucine-rich repeat protein LRP - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
 C:Accession: T07079
 R:Tornero, P.; Mayda, E.; Gomez, M.D.; Canas, L.; Conejero, V.; Vera, P.
 Plant J. 10, 315-330, 1996
 A:Title: Characterization of LRP, a leucine-rich repeat (LRR) protein from tomato pla
 A:Reference number: 215903; MUID:96367673
 A:Accession: T07079
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-221 <TOR>
 A:Cross-references: EMBL:X95269; NID:g1619299; PIDN:CA64565.1; PID:g1619300
 A:Experimental source: cultivar VFN8; leaf
 C:Genetics:
 A:Gene: LRP
 A:introns: 78/2; 102/2; 150/2; 174/2; 198/2

Query Match 0.8%; Score 8; DB 2; Length 221;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 758 SIGNLKNL 765
 |||||||
 Db 140 SIGNLKNL 147

RESULT 9

ADPSGP
 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) [validated] - Pseudomonas pu
 N:Alternate names: phospho-2-dehydro-3-deoxygluconate aldolase; phospho-2-keto-3-deox
 C:Species: Pseudomonas putida
 C:Date: 30-Nov-1980 #sequence_revision 30-Nov-1980 #text_change 26-May-2000
 C:Accession: A01105
 R:Suzuki, N.; Wood, W.A.
 J. Biol. Chem. 255, 3427-3435, 1980
 A:Title: Complete primary structure of 2-keto-3-deoxy-6-phosphogluconate aldolase.
 A:Reference number: A92273; MUID:80159956
 A:Accession: A01105
 A:Molecule type: protein
 A:Residues: 1-225 <SUZ>
 R:Maavidis, I.M.; Tullinsky, A.
 Biochemistry 15, 4410-4417, 1976
 A:Title: The folding and quaternary structure of trimeric 2-keto-3-deoxy-6-phosphoglu
 A:Reference number: A90397; MUID:77022062
 A:Contents: annotation; X-ray crystallography; 3.5 angstroms
 C:Superfamily: 2-dehydro-3-deoxyphosphogluconate aldolase
 C:Keywords: aldehyde-lyase; carbon-carbon lyase; homotrimer
 F:56/60/63/Active site: Glu, Arg, His status predicted
 F:144/Active site: Lys (covalent pyruvate-binding) #status experimental

Query Match 0.8%; Score 8; DB 1; Length 225;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 660 RTLEVTLR 667
|||||
Db 53 RTLEVTLR 60

RESULT 10
C66317
protein T10022.23 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C66317
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huiziar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C66317
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <STO>
A:Cross-references: GB:AE005172; NID:q8671774; PIDN:AAF78380.1; GSPDB:GN00141
C:Genetics:
A:Gene: T10022.23
A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 MAMLLKLR 240
|||||
Db 226 MAMLLKLR 233

RESULT 11
T28975
hypothetical protein T28A11.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T28975
R:Roiffin, T.
submitted to the EMBL Data Library, January 1997
A:Description: The sequence of C. elegans cosmid T28A11.
A:Reference number: Z20550
A:Accession: T28975
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-350 <ROH>
A:Cross-references: EMBL:U80027; PIDN:AAC48131.1; GSPDB:GN00023; CESP:T28A11.10
A:Experimental source: strain Bristol N2; clone T28A11
C:Genetics:
A:Gene: CESP:T28A11.10
A:Map position: 5
A:Introns: 64/3; 120/2; 149/3; 188/3; 222/3; 254/3

Query Match 0.8%; Score 8; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 686 SATSLRLQ 693
|||||
Db 237 SATSLRLQ 244

RESULT 12
E81269
hypothetical protein Cj1713 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: E81269
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: E81269
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <PAR>
A:Cross-references: GB:AL139079; GB:AL111168; NID:q6968971; PIDN:CAB73699.1; PID:q696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1713
C:Superfamily: conserved hypothetical protein H10365

Query Match 0.8%; Score 8; DB 2; Length 356;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AVKILQON 843
|||||
Db 184 AVKILQON 191

RESULT 13
F81816
phosphoserine transaminase (EC 2.6.1.52) NMA1894 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: F81816
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jorgels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: A81775; MUID:20222556
A:Accession: F81816
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:q7380371; PIDN:CAB85115.1; PID:q738
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: serC; NMA1894
C:Superfamily: phosphoserine aminotransferase
C:Keywords: aminotransferase

Query Match 0.8%; Score 8; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 RUTREIR 925
|||||
Db 114 RUTREIR 121

RESULT 14
H81059
phosphoserine aminotransferase NMB1640 [imported] - Neisseria meningitidis (strain MC
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81059
R:Reitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rl, H.; Qin, H.; Yamahevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Piza, M.

Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vg
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: H81059
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <TET>
A:Cross-references: GB:AE002514; GB:AE002098; NID:g7226886; PIDN:AAF41989.1; PID:g722688
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1640
C:Superfamily: phosphoserine aminotransferase

Query Match 0.8%; Score 8; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 RLTDTETR 925
DB 114 RLTDTETR 121

RESULT 15

GRECY

tyrosine-specific transport protein - *Escherichia coli*

N:Alternate names: tyrosine permease

C:Species: *Escherichia coli*

C>Date: 30-Jun-1990 #sequence_revision 31-Oct-1997 #text_change 16-Jul-1999

C:Accession: C64954; JS0146

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Ch

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: C64954

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-403 <BLAT>

A:Cross-references: GB:AE000284; GB:U00096; NID:g1788214; PIDN:AAC74977.1; PID:g1788218;

A:Experimental source: strain K-12, substrain MG1655

J:Wooley, P.J.; Pittard, A.J.

R:Microbiol. 170, 4946-4949, 1988

A:Title: DNA sequence of the gene (tyrP) encoding the tyrosine-specific transport system

A:Reference number: JS0146; MUID:89008121

A:Accession: JS0146

A:Molecule type: DNA

A:Residues: 1-130, 'RRVAVL', 136-221, 'ECD', 225-403 <MOO>

A:Cross-references: GB:M23240; NID:g148088; PIDN:AAA24705.1; PID:g148089

C:Genetics:

A:Gene: tyrP

A:Map position: 42 min

A:Start codon: GTG

C:Function:

A:Note: involved in transporting tyrosine across the cytoplasmic membrane

C:Superfamily: tyrosine-specific transport protein

C:Keywords: amino acid transport; inner membrane; Transmembrane protein; transport prote

F:8-24/Domain: transmembrane #status predicted <TM1>

F:37-53/Domain: transmembrane #status predicted <TM2>

F:122-138/Domain: transmembrane #status predicted <TM3>

F:149-165/Domain: transmembrane #status predicted <TM4>

F:216-232/Domain: transmembrane #status predicted <TM5>

F:277-293/Domain: transmembrane #status predicted <TM6>

F:335-351/Domain: transmembrane #status predicted <TM7>

F:378-394/Domain: transmembrane #status predicted <TM8>

Query Match 0.8%; Score 8; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 LINGLOAL 161
DB 255 LINGLOAL 262

RESULT 16

D85804

tyrosine-specific transport system [imported] - *Escherichia coli* (strain O157:H7)C:Species: *Escherichia coli*

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: D85804

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85804

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-403 <STO>

A:Cross-references: GB:AE005174; NID:g12515873; PIDN:AAG56896.1; GSPDB:GN00145; UMGF:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: tyrP

C:Superfamily: tyrosine-specific transport protein

Query Match 0.8%; Score 8; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 LINGLOAL 161
DB 255 LINGLOAL 262

RESULT 17

B70664

hypothetical protein RV1842c - *Mycobacterium tuberculosis* (strain H37RV)C:Species: *Mycobacterium tuberculosis*

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70664

R:Conor, R.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: B70664

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-455 <COL>

A:Cross-references: GB:Z83859; GB:AL123456; NID:g3261678; PIDN:CAB06112.1; PID:g17812

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV1842c

C:Superfamily: hypothetical protein HI0107

Query Match 0.8%; Score 8; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 QFGALTAE 296
DB 214 QFGALTAE 221

RESULT 18

JH0197

muscarinic acetylcholine receptor M2 - rat

C:Species: *Rattus norvegicus* (Norway rat)

C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 24-Nov-1999
C:Accession: JH0197; D37121
R:Lat, J.; Bloom, J.W.; Yamamura, H.I.; Roeske, W.R.
Life Sci. 47, 1001-1013, 1990
A:Title: Amplification of the rat m2 muscarinic receptor gene by the polymerase chain re
A:Reference number: JH0197; MUID:91041524
A:Accession: JH0197
A:Molecule type: DNA
A:Residues: 1-466 <LA1>
R:Kutenbach, E.; Curtis, C.A.M.; Pedder, E.K.; Aitken, A.; Harris, A.C.M.; Hulme, E.C.
J. Biol. Chem. 265, 13702-13708, 1990
A:Title: Muscarinic acetylcholine receptors. Peptide sequencing identifies residues inv
A:Reference number: A37121; MUID:90337982
A:Accession: D37121
A:Status: preliminary
A:Molecule type: protein
A:Residues: 60-122 <KUR>
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; transme
F:23-48/Domain: transmembrane #status predicted <TM1>
F:61-85/Domain: transmembrane #status predicted <TM2>
F:98-119/Domain: transmembrane #status predicted <TM3>
F:139-162/Domain: transmembrane #status predicted <TM4>
F:184-207/Domain: transmembrane #status predicted <TM5>
F:389-409/Domain: transmembrane #status predicted <TM6>
F:421-442/Domain: transmembrane #status predicted <TM7>

Query Match 0.8%; Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSLV 707
|||||||
Db 29 VAGSLSLV 36

RESULT 19
S10126
muscarinic acetylcholine receptor M2 - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 24-Nov-1999
C:Accession: S10126
R:Peralta, E.G.; Ashkenazi, A.; Winslow, J.W.; Smith, D.H.; Ramachandran, J.; Capon, D.J.
EMBO J. 6, 3923-3929, 1987
A:Title: Distinct primary structures, ligand-binding properties and tissue-specific exp
A:Reference number: S04326; MUID:88166632
A:Accession: S10126
A:Molecule type: DNA
A:Residues: 1-466 <PER>
A:Cross-references: EMBL:X15264; NID:g32319; PIDN:CAA3335.1; PID:g32320
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phospho
F:23-48/Domain: transmembrane #status predicted <TM1>
F:60-89/Domain: transmembrane #status predicted <TM2>
F:98-119/Domain: transmembrane #status predicted <TM3>
F:139-162/Domain: transmembrane #status predicted <TM4>
F:184-207/Domain: transmembrane #status predicted <TM5>
F:389-409/Domain: transmembrane #status predicted <TM6>
F:421-442/Domain: transmembrane #status predicted <TM7>
F:2,6,9/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSLV 707
|||||||
Db 29 VAGSLSLV 36

RESULT 20

S10856
muscarinic acetylcholine receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Nov-1999
C:Accession: S10856
R:Goodyne, J.; Robinson, D.A.; Fitzgerald, M.G.; Chung, F.Z.; Kerlavage, A.R.; Lentes
Proc. Natl. Acad. Sci. U.S.A. 84, 8296-8300, 1987
A:Title: Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic
A:Reference number: S10855; MUID:88068581
A:Accession: S10856
A:Molecule type: mRNA
A:Residues: 1-466 <GOC>
A:Cross-references: EMBL:J03025; NID:g203461; PIDN:AAA0926.1; PID:g203462
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phos
F:23-48/Domain: transmembrane #status predicted <TM1>
F:61-85/Domain: transmembrane #status predicted <TM2>
F:98-119/Domain: transmembrane #status predicted <TM3>
F:139-162/Domain: transmembrane #status predicted <TM4>
F:184-207/Domain: transmembrane #status predicted <TM5>
F:389-409/Domain: transmembrane #status predicted <TM6>
F:421-442/Domain: transmembrane #status predicted <TM7>

Query Match 0.8%; Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 VAGSLSLV 707
|||||||
Db 29 VAGSLSLV 36

RESULT 21
A27386
muscarinic acetylcholine receptor, cardiac - pig
N:Alternate names: muscarinic acetylcholine receptor M2
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 31-Dec-2000
C:Accession: A27386; A25656
R:Peralta, E.G.; Winslow, J.W.; Peterson, G.L.; Smith, D.H.; Ashkenazi, A.; Ramachand
Science 236, 600-605, 1987
A:Title: Primary structure and biochemical properties of an M-2 muscarinic receptor.
A:Reference number: A27386; MUID:87206169
A:Accession: A27386
A:Molecule type: DNA; mRNA
A:Residues: 1-466 <PER>
A:Cross-references: GB:M6331; NID:g164311; PIDN:AA30986.1; PID:g164313
A:Experimental source: atrial muscle
A>Note: the protein sequence derived from the mRNA clones differs from that of the ge
R:Castro, J.
FEBS Lett. 209, 367-372, 1986
A:Title: Primary structure of porcine cardiac muscarinic acetylcholine receptor deduc
A:Reference number: A25656; MUID:87080790
A:Accession: A25656
A:Molecule type: mRNA
A:Residues: 1-329, 'K', 331-466 <RUB>
A:Cross-references: GB:X04708; NID:g1859; PIDN:CAA28413.1; PID:g1860
C:Superfamily: vertebrate rhodopsin
C:Keywords: cardiac muscle; G protein-coupled receptor; glycoprotein; heart; neurotra
F:23-48/Domain: transmembrane #status predicted <TM1>
F:61-85/Domain: transmembrane #status predicted <TM2>
F:98-119/Domain: transmembrane #status predicted <TM3>
F:139-162/Domain: transmembrane #status predicted <TM4>
F:184-207/Domain: transmembrane #status predicted <TM5>
F:389-409/Domain: transmembrane #status predicted <TM6>
F:421-442/Domain: transmembrane #status predicted <TM7>

Query Match 0.8%; Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 VAGSLSLV 707
|||||
Db 29 VAGSLSLV 36

RESULT 22

157961

glucuronosyltransferase (EC 2.4.1.17) precursor - rat

N:Alternate names: glucuronosyltransferase 1 B1; morphine UGT

C:Species: Rattus norvegicus (Norway rat)

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 29-Sep-1999

C:Accession: 157961; S51197; S68333

R:Coffman, B.L.; Green, M.D.; King, C.D.; Tephy, T.R.

Mol. Pharmacol. 47, 1101-1105, 1995

A:Title: Cloning and stable expression of a cDNA encoding a rat liver UDP-glucuronosyltr

A:Reference number: 157961; MUID:95327065

A:Accession: 157961

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-535 <RES>

A:Cross-references: EMBL:U20551; NID:g695161; PIDN:AAC52219.1; PID:g695162

R:Ishii, Y.; Tsuruda, K.; Tanaka, M.; Oguri, K.

Arch. Biochem. Biophys. 315, 345-351, 1994

A:Title: Purification of a phenobarbital-inducible morphine UDP-glucuronosyltransferase is

A:Reference number: S51197; MUID:95077409

A:Accession: S51197

A:Molecule type: protein

A:Residues: 30-41 <ISH>

R:Ikuhiro, S.; Eml, Y.; Iyanagi, T.

Arch. Biochem. Biophys. 324, 267-272, 1995

A:Title: Identification and analysis of drug-responsive expression of UDP-glucuronosyltr

A:Reference number: S68333; MUID:96132654

A:Accession: S68333

A:Molecule type: protein

A:Residues: 30-37 <IKU>

C:Genetics:

A:Gene: UGT1.1

C:Superfamily: glucuronosyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 570 ALSOEFEA 577
|||||
Db 287 ALSOEFEA 294

RESULT 23

E85431
hypothetical protein AT4g36550 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: E85431

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: E85431

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-660 <STO>

A:Cross-references: GB:NC_001268; NID:g7270603; PIDN:CAB80321.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g36550

A:Map position: 4

Query Match 0.8%; Score 8; DB 2; Length 660;
Best Local Similarity 100.0%; Pred. No. 29;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 894 GSLSLKL 901
|||||
Db 465 GSLSLKL 472

RESULT 24

T18840

hypothetical protein C01G6.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T18840

R:Berts, M.

submitted to the EMBL Data Library, August 1994

A:Reference number: Z19029

A:Accession: T18840

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-806 <WIL>

A:Cross-references: EMBL:Z55955; PIDN:CAA84639.1; GSPDB:GN00020; CESP:C01G6.8

A:Experimental source: clone C01G6

C:Genetics:

A:Gene: CESP:C01G6.8

A:Map position: 2

A:Introns: 36/3; 170/3; 217/3; 636/3; 760/1

Query Match 0.8%; Score 8; DB 2; Length 806;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 713 NITSLWE 720
|||||
Db 709 NITSLWE 716

RESULT 25

T10504

disease resistance protein Cf-2.1 - current tomato

C:Species: Lycopersicon pimpinellifolium (current tomato)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000

C:Accession: T10504; T10515

R:Dixon, M.S.; Jones, D.A.; Keddie, J.S.; Thomas, C.M.; Harrison, K.; Jones, J.D.

Cell 84, 451-459, 1996

A:Title: The tomato Cf-2 disease resistance locus comprises two functional genes enco

A:Reference number: Z17062; MUID:96190812

A:Accession: T10504

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1112 <DIY>

A:Cross-references: EMBL:U42444; NID:g1184074; PIDN:AAC15779.1; PID:g1184075

A:Experimental source: cultivar Cf 2

A:Accession: T10515

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1066, 'I', 1068-1085, 'E', 1087-1110, 'R', 1112 <DI2>

A:Cross-references: EMBL:U42445; NID:g1184076; PIDN:AAC15780.1; PID:g1184077

A:Experimental source: cultivar Cf 2

Query Match 0.8%; Score 8; DB 2; Length 1112;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 SLGNLKL 765
|||||
Db 330 SLGNLKL 337

RESULT 26

T43148

probable protein-tyrosine-phosphatase (EC 3.1.3.48) - horn shark

N:Alternate names: CD45 homolog
 C:Species: Heterodontus francisci (horn shark)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
 C:Accession: T43148
 R:Okumura, M., Matthews, R.J., Robb, B., Bork, P., Thomas, M.L.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: 222317
 A:Accession: T43148
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1200 <OKU>
 A:Cross-references: EMBL:U34750; NID:g1304393; PID:g1335805; PIDN:AAE01087.1
 C:Superfamily: leukocyte common antigen; leukocyte common antigen cytosolic domain homol
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatas

Query Match 0.8%; Score 8; DB 2; Length 1200;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 236 LTKLRQRV 243
 |||||||
 DB 729 LTKLRQRV 736

RESULT 27
 E70763
 hypothetical protein Rv1560 - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: E70763
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: E70763
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-72 <COL>
 A:Cross-references: GB:274020; GB:AL123456; NID:g3261584; PIDN:CAA98331.1; PID:e248537;
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv1560

Query Match 0.7%; Score 7; DB 2; Length 72;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 EGDLDL 101
 |||||||
 DB 60 EGDLDL 66

RESULT 28
 E65027
 hypothetical protein b2510 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
 C:Accession: E65027
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
 Science 277, 1453-1462, 1997
 A:Rose, D.J.; Mau, B.; Shao, Y.
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65027
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-83 <BLAT>
 A:Cross-references: GB:AE000337; GB:U00096; NID:g1788850; PIDN:AAC75563.1; PID:g1788857;

A:Experimental source: strain K-12, substrain MG1655

Query Match 0.7%; Score 7; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 631 TGGIHME 637
 |||||||
 DB 8 TGGIHME 14

RESULT 29
 S35715
 glycine-rich protein (clone DC 7.1), embryonic - carrot
 C:Species: Daucus carota (carrot)
 C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 11-Jan-2000
 C:Accession: S35715; S08024
 R:Alteith, F.; Richter, G.
 Planta 183, 17-24, 1990
 A:Title: Gene expression during induction of somatic embryogenesis in carrot cell sus
 A:Reference number: S35714
 A:Accession: S35715
 A:Molecule type: mRNA
 A:Residues: 1-96 <ALE>
 A:Cross-references: EMBL:X15706; NID:g18348; PID:g18349
 C:Superfamily: Arabidopsis glycine-rich protein 3
 C:Keywords: embryo; transmembrane protein
 F:5-25/Domain: transmembrane #status predicted <TMM>
 F:40-72/Region: glycine/histidine-rich

Query Match 0.7%; Score 7; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 518 LGLSLTA 524
 |||||||
 DB 8 LGLSLTA 14

RESULT 30
 H71630
 hypothetical protein RP709 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: H71630
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499
 A:Accession: H71630
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-98 <AND>
 A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA15144.1; PID:e134
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: RP709

Query Match 0.7%; Score 7; DB 2; Length 98;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 625 EKAEDT 631
 |||||||
 DB 36 EKAEDT 42

RESULT 31
 D72710
 hypothetical protein APE1099 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: D72710
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; H aikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
 A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
 A:Reference number: A72450; MWID:99310339
 A:Accession: D72710
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <RAW>
 A:Cross-references: DBJ:AF000060; NID:95104188; PIDN:BA80084.1; PID:d1043870; PID:9510
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1099

Query Match 0.7%; Score 7; DB 2; Length 103;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 518 LUGLSIA 524
 |||||
 Db 53 LUGLSIA 59

RESULT 32
 S04103
 hemolysin A - Morganella morganii (fragment)
 C:Species: Morganella morganii
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 15-Nov-1996
 C:Accession: S04103
 R:Koronakis, V.; Koronakis, E.; Hughes, C.
 EMBO J. 8, 595-605, 1989
 A>Title: Isolation and analysis of the C-terminal signal directing export of Escherichia
 A:Reference number: S04101; MWID:89251588
 A:Accession: S04103
 A>Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-104 <KOR>
 C:Genetics:
 A:Gene: hlyA
 C:superfamily: hemolysin A; hemolysin A homology
 C:Keywords: calcium binding; cytotoxins; exotoxin; hemolysis; lipoprotein; tandem repeat

Query Match 0.7%; Score 7; DB 2; Length 104;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 456 SSLLTSH 462
 |||||
 Db 80 SSLLTSH 86

RESULT 33
 S57388
 hypothetical protein orf 00958 - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C>Date: 20-Oct-1995 #sequence_revision 05-Sep-1996 #text_change 20-Oct-2000
 C:Accession: S57388; S50423
 R:Zumstern, E.; Pearson, B.M.; Kalogeropoulos, A.; Schweitzer, M.
 Yeast 11, 975-986, 1995
 A>Title: A 29,425 kb segment on the left arm of yeast chromosome XV contains more than 1
 A:Reference number: S57374; MWID:96021609
 A:Accession: S57388
 A>Status: nucleic acid sequence not shown; conceptual translation of pseudogene
 A:Molecule type: DNA
 A:Residues: 1-107 <ZUM>
 A:Cross-references: EMBL:X83121; NID:9600461; PIDN:CAA58198.1; PID:9600477
 C:Comment: There is no evidence that this sequence is expressed.
 C:Genetics:

A:Map position: 15L
 C:Keywords: pseudogene
 Query Match 0.7%; Score 7; DB 4; Length 107;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 455 LSSLLTS 461
 |||||
 Db 42 LSSLLTS 48

RESULT 34
 B82830
 hypothetical protein XF0247 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: B82830
 R:anonymous, The Xylella fastidiosa Consortium of the organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MWID:20365717
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: B82830
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-111 <SIM>
 A:Cross-references: GB:AE003878; GB:AE003849; NID:99105052; PIDN:AAF83060.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Agencio, M.; Alvarenga, R.
 Britone, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carer
 as-Neto, E.; Docena, C.; El-Dorri, H.; Faciniani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J.; de M., de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tshukako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF0247

Query Match 0.7%; Score 7; DB 2; Length 111;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 9 PALIORM 15
 |||||
 Db 15 PALIORM 21

RESULT 35
 S38608
 ribosomal protein s19 - euglenid (Astasia longa) plastid
 C:Species: plastid Astasia longa
 C>Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 13-Aug-1999
 C:Accession: S38608
 R:Gockel, G.; Baler, S.; Hachtel, W.
 submitted to the EMBL Data Library, November 1993
 A:Reference number: S38590
 A:Accession: S38608
 A:Molecule type: DNA
 A:Residues: 1-118 <GOC>
 A:Cross-references: EMBL:X75653; NID:9414863; PIDN:CAA53332.1; PID:9414874
 C:Genetics:
 A:Gene: rps19
 A:Genome: plastid

A:Introns: 26/3; 69/3
C:Superfamily: Escherichia coli ribosomal protein s19
C:Keywords: chloroplast; plastid; protein biosynthesis; ribosome

Query Match 0.7%; Score 7; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 963 NLKOLVF 969
|||||
DB 26 NLKOLVF 32

RESULT 36
A29774
T-cell receptor alpha chain precursor V region (HAPO1) - human
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-May-1997
C:Accession: A29774
R:Yoshikai, Y.; Kimura, N.; Toyonaga, B.; Mak, T.W.
J. Exp. Med. 164, 90-103, 1986
A:Title: Sequences and repertoire of human T cell receptor alpha chain variable region
A:Reference number: A92778; MUID:86253078
A:Accession: A29774
A:Molecule type: mRNA
A:Residues: 1-133 <YOS>
C:Genetics:
A:Map position: 14q11.2
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: T-cell receptor

Query Match 0.7%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 895 SLSLKLK 901
|||||
DB 2 SLSLKLK 8

RESULT 37
D72539
hypothetical protein APE1601 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72539
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kawa Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339
A:Accession: D72539
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <KAW>
A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAAB0601.1; PID:d1044387; PID:g510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1601

Query Match 0.7%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 686 SATSLRL 692
|||||
DB 14 SATSLRL 20

RESULT 38

S04069
glycine-rich protein - red goosefoot
C:Species: Chenopodium rubrum (red goosefoot)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C:Accession: S04069
R:Kaidemoff, R.; Richter, G.
Nucleic Acids Res. 17, 2853, 1989
A:Title: Sequence of cDNA for a novel light-induced glycine-rich protein.
A:Reference number: S04069; MUID:89240041
A:Accession: S04069
A:Molecule type: mRNA
A:Residues: 1-144 <KAI>
A:Cross-references: EMBL:X14067; NID:g18147; PIDN:CAA32230.1; PID:g18148
A:Note: the authors translated the codon ATC for residue 5 as Asn
C:Superfamily: Arabidopsis glycine-rich protein 3
C:Keywords: transmembrane protein

Query Match 0.7%; Score 7; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 518 LLGLSIA 524
|||||
DB 8 LLGLSIA 14

RESULT 39
S35716
glycine-rich protein (clone DC 9.1), embryonic - carrot
C:Species: Daucus carota (carrot)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 24-Nov-1999
C:Accession: S35716
R:Alaith, F.; Richter, G.
Planta 183, 17-24, 1990
A:Title: Gene expression during induction of somatic embryogenesis in carrot cell sus
A:Reference number: S35714
A:Accession: S35716
A:Molecule type: mRNA
A:Residues: 1-144 <ALE>
C:Superfamily: Arabidopsis glycine-rich protein 3
C:Keywords: embryo; transmembrane protein
E:5-25/Domain: transmembrane #status predicted <TM>
F:37-107/Region: 6-residue repeat

Query Match 0.7%; Score 7; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 518 LLGLSIA 524
|||||
DB 8 LLGLSIA 14

RESULT 40
T48221
hypothetical protein T7H20.50 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48221
R:Bevan, M.; Peters, S.A.; Van Staveren, M.; Dirkse, W.; Slietema, W.; Bancroft, I.; submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24488
A:Accession: T48221
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BEV>
A:Cross-references: EMBL:AL162508
A:Experimental source: cultivar Columbia; BAC clone T7H20
C:Genetics:
A:Map position: 5
A:Note: T7H20.50

Query Match 0.7%: Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 IFSSATS 689
 |||||
 DB 54 IFSSATS 60

RESULT 41
 H83076
 Probable peptidyl-prolyl cis-trans isomerase, Fkbp-type PA4558 [imported] - Pseudomonas
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83076
 R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yun, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337
 A:Accession: H83076
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <SPU>
 A:Cross-references: GB:AE00486; GB:AE004091; NID:g9950793; PIDN:AG07946.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4558

Query Match 0.7%: Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 ELAEGIL 322
 |||||
 DB 92 ELAEGIL 98

RESULT 42
 S25499
 T-cell receptor alpha chain - human
 C:Species: Homo sapiens (man)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
 C:Accession: S25499
 R:Spurkland, A.; Gedde-Dahl, T.; Hansen, T.; Vartdal, F.; Gaudernack, G.; Thorsby, E.
 submitted to the EMBL Data Library, August 1992
 A:Description: T cells specific for a given peptide presented by the same HLA-DQ molecu
 A:Reference number: S25499
 A:Accession: S25499
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-147 <SPU>
 A:Cross-references: EMBL:Z14996; NID:g36170; PIDN:CAAT8717.1; PID:g36171
 C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: T-cell receptor

Query Match 0.7%: Score 7; DB 2; Length 147;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 895 SLSSILK 901
 |||||
 DB 3 SLSSILK 9

RESULT 43
 B82867
 hypothetical protein xfa0051 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: B82867
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: B82867
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-157 <SIM>
 A:Cross-references: GB:AE003851; NID:g9112238; PIDN:AAF85619.1; GSPDB:GN00130; XFSC:X
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramag, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: Xfa0051
 A:Genome: plasmid
 A:Note: plasmid pXF5.1

Query Match 0.7%: Score 7; DB 2; Length 157;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 SLSLVLS 709
 |||||
 DB 113 SLSLVLS 119

RESULT 44
 MORFL2
 myosin I2 (DTNB) regulatory light chain precursor, skeletal muscle - rat
 N:Alternate names: MLC2 light chain; myosin g2 chain
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999
 C:Accession: A03041; I55219
 R:Nudel, U.; Calvo, J.M.; Shan, M.; Levy, Z.
 Nucleic Acids Res. 12, 7175-7186, 1984
 A:Title: The nucleotide sequence of a rat myosin light chain 2 gene.
 A:Reference number: A03041; MUID:85014159
 A:Accession: A03041
 A:Molecule type: DNA
 A:Residues: 1-169 <NUD>
 A:Cross-references: GB:X00975; NID:g56726; PIDN:CAA25480.1; PID:g825539
 R:Garfinkel, L.I.; Periasamy, M.; Nadal-Ginard, B.
 J. Biol. Chem. 257, 11078-11086, 1982
 A:Title: Cloning and characterization of cDNA sequences corresponding to myosin light
 A:Reference number: I55219; MUID:82265830
 A:Accession: I55219
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 70-169 <RES>
 A:Cross-references: GB:J00754; NID:g205600; PIDN:AAA41660.1; PID:g205601
 C:Comment: Adult rat skeletal muscle contains three light chains: MLC1, MLC2, and MLC
 C:Genetics:
 A:Gene: MLC2
 A:Introns: 1/3; 32/3; 58/1; 93/1; 119/2; 135/3
 C:Superfamily: calmodulin; calmodulin repeat homology
 C:Keywords: calcium binding; EF hand; muscle contraction; phosphoprotein

F:2-169/Product: myosin I2 (DTNB) regulatory light chain #status predicted <MAT>
F:25-57/Domain: calmodulin repeat homology <EF1>
F:95-127/Domain: calmodulin repeat homology <EF1>
F:16/Binding site: phosphate (Ser) (covalent) #status predicted
F:36,40,42,44,49/Binding site: calcium (Asp, Asn, Asp, Ile, Asp) #status predicted

Query Match 0.7%; Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1014 VITGAFK 1020
Db 99 VITGAFK 105

RESULT 45

MORBU
myosin I2 (DTNB) regulatory light chain, skeletal muscle - rabbit
N:Alternate names: MLC2 regulatory light chain; myosin 92 chain
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-May-1979 #sequence_revision 01-Mar-1996 #text_change 24-Nov-1999
C:Accession: S12691; A03040; S13445; I46493; I46494
R:Maeda, K.; Mueller-gerhardt, E.; Wittinghofer, A.
Nucleic Acids Res. 18, 6687, 1990
A:Title: Sequence of two isoforms of myosin light chain 2 isolated from a rabbit fast SK
A:Reference number: S12691; MUID:91067462
A:Accession: S12691
A:Molecule type: mRNA
A:Residues: 1-170 <MAE>
A:Cross-references: EMBL:X54043; NID:q1636; PIDN:CAA37976.1; PID:q1637
A>Note: the authors translated the codon TTC for residue 20 as Ser and TGC for residue 1
R:Matsuda, G.; Malta, T.; Suzuyama, Y.; Setoguchi, M.; Umegane, T.
Hoppe-Seyler's Z. Physiol. Chem. 359, 629-640, 1978
A:Title: The amino acid sequences of the tryptic, chymotryptic and peptic peptides from
A:Reference number: A03040; MUID:78216701
A:Accession: A03040
A:Molecule type: protein
A:Residues: 3-170 <MAT>
R:Matsuda, G.; Malta, T.; Suzuyama, Y.; Setoguchi, M.; Umegane, T.
J. Biochem. 81, 809-811, 1977
A:Title: Amino acid sequence of the L-2 light chain of rabbit skeletal muscle myosin.
A:Reference number: S13445; MUID:77187770
A:Accession: S13445
A:Status: preliminary
A:Molecule type: protein
A:Residues: 3-170 <MA2>
R:Putney, S.D.; Herlihy, W.C.; Schimmel, P.
Nature 302, 718-721, 1983
A:Title: A new tropomyosin T and cDNA clones for 13 different muscle proteins, found by shc
A:Reference number: I46471; MUID:83167564
A:Accession: I46493
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 28-39 <PUT>
A:Cross-references: EMBL:V00887; NID:q1629; PIDN:CAA24255.1; PID:q929758
A:Accession: I46494
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 113-146 <PU2>
A:Cross-references: EMBL:V00888; NID:q1631; PIDN:CAA24256.1; PID:q929759
C:Superfamily: calmodulin; calmodulin repeat homology
C:Keywords: blocked amino end; calcium binding; EF hand; muscle; muscle contraction; phc
F:26-58/Domain: calmodulin repeat homology <EF1>
F:96-128/Domain: calmodulin repeat homology <EF2>
F:12/Modified site: blocked amino end (Ala) (in mature form) (probably trimethylated) #st
F:17/Binding site: phosphate (Ser) (covalent) #status predicted
F:39,41,43,45,50/Binding site: calcium (Asp, Asn, Asp, Ile, Asp) #status predicted

Query Match 0.7%; Score 7; DB 1; Length 170;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 VITGAFK 1020
Db 100 VITGAFK 106

RESULT 46

myosin I2 regulatory light chain type 1, skeletal muscle - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 13-Aug-1999
C:Accession: S12855
R:Maeda, K.; Mueller-gerhardt, E.; Wittinghofer, A.
Nucleic Acids Res. 18, 6687, 1990
A:Title: Sequence of two isoforms of myosin light chain 2 isolated from a rabbit fast
A:Reference number: S12691; MUID:91067462
A:Accession: S12855
A:Molecule type: mRNA
A:Residues: 1-170 <MAE>
A:Cross-references: EMBL:X54042; NID:q1634; PIDN:CAA37975.1; PID:q1635
A>Note: the authors translated the codon TTC for residue 20 as Ser and TGC for residu
C:Superfamily: calmodulin; calmodulin repeat homology
C:Keywords: calcium binding; EF hand; muscle; muscle contraction; phosphoprotein; ske
F:26-58/Domain: calmodulin repeat homology <EF1>
F:96-128/Domain: calmodulin repeat homology <EF2>

Query Match 0.7%; Score 7; DB 2; Length 170;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1014 VITGAFK 1020
Db 100 VITGAFK 106

RESULT 47

JH0783
diamine N-acetyltransferase (EC 2.3.1.57) - human
N:Alternate names: protein DKFZ586G1923.1; spermidine/spermine N1-acetyltransferase
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-May-2000
C:Accession: JH0783; A39082; J01169; S61530; T08804; S31609
R:Xiao, L.; Celano, P.; Mank, A.R.; Griffin, C.; Jabs, E.W.; Hawkins, A.L.; Casero Jr
Biochem. Biophys. Res. Commun. 187, 1493-1502, 1992
A:Title: Structure of the human spermidine/spermine N1-acetyltransferase gene.
A:Reference number: JH0783; MUID:93038627
A:Accession: JH0783
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-171 <XIA>
A:Cross-references: EMBL:214136; NID:q36606; PIDN:CAA78509.1; PID:q36607
R:Casero Jr., R.A.; Celano, P.; Eryin, S.J.; Applegreen, N.B.; Wiest, L.; Pegg, A.E.
J. Biol. Chem. 266, 810-814, 1991
A:Title: Isolation and characterization of a cDNA clone that codes for human spermidl
A:Reference number: A39082; MUID:91093277
A:Accession: A39082
A:Molecule type: mRNA
A:Residues: 1-25, 'K', '27-171 <CAS>
A:Cross-references: GB:M5580; NID:q338335; PIDN:AAA63260.1; PID:q338336
R:Xiao, L.; Celano, P.; Mank, A.R.; Pegg, A.E.; Casero Jr., R.A.
Biochem. Biophys. Res. Commun. 179, 407-415, 1991
A:Title: Characterization of a full-length cDNA which codes for the human spermidine/
A:Reference number: J01169; MUID:91354284
A:Accession: J01169
A:Molecule type: mRNA
A:Residues: 1-25, 'K', '27-171 <XIA>
A:Cross-references: GB:M77693; NID:q338391; PIDN:AAA60573.1; PID:q338392
A:Experimental source: lung carcinoma cell line NCI H157
R:Xiao, L.; Casero Jr., R.A.
Biochem. J. 313, 691-696, 1996
A:Title: Differential transcription of the human spermidine/spermine N(1)-acetyltrans
A:Reference number: S61530; MUID:96152560

A:Accession: S61530
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-171 <X13>
A:Cross-references: EMBL:U0369; NID:q1103903; PIDN:AAA98854.1; PID:q1103904
R:Ansgorge, W.; Wilner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216472
A:Accession: T08804
A:Molecule type: mRNA
A:Residues: 76-171 <ANS>
A:Cross-references: EMBL:AL050290
A:Experimental source: adult uterus; clone DKFZp586g1923
C:Genetics:
A:Gene: GDB:SAT
A:Cross-references: GDB:127512; OMIM:313020
A:Map position: XP22.1-XP22.1
A:Introns: 22/3; 40/1; 68/1; 102/1; 115/3
A:Note: DKFZp586g1923.1
C:Function:
A:Description: responsible for catabolism of cellular polyamines
C:Keywords: acetyltransferase; coenzyme A; polyamine biosynthesis
F:1-171/Product: diamine N-acetyltransferase #status predicted <MAT1>
F:76-171/Product: diamine N-acetyltransferase, splice variant #status predicted <MAT2>

Query Match 0.7%; Score 7; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 313 LKELAE 319
|||||
DB 20 LKELAE 26

RESULT 48
S25724
hypothetical protein 1 (phbc 5' region) - Rhodococcus ruber
N:Alternate names: PHA-synthase
C:Species: Rhodococcus ruber
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 22-Oct-1999
C:Accession: S25724
R:Pieper, U.; Steinhuechel, A.
FEBS Microbiol. Lett. 96, 73-80, 1992
A:Title: Identification, cloning and sequence analysis of the poly(3-hydroxyalkanoic acid)
A:Reference number: S25724
A:Accession: S25724
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <PIE>
A:Cross-references: EMBL:X66407; NID:g46398; PIDN:CAA47034.1; PID:g46399

Query Match 0.7%; Score 7; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 292 ALTAEVG 298
|||||
DB 112 ALTAEVG 118

RESULT 49
E69411
conserved hypothetical protein AF1294 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
C:Accession: E69411
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weisman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Springs, T.; Arlrich, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Moese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: E69411
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-178 <KLE>
A:Cross-references: GB:AE001015; GB:AE000782; NID:g2689338; PIDN:AA89954.1; PID:g264
C:Superfamily: conserved hypothetical protein MJ0645

Query Match 0.7%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 290 FGALTAE 296
|||||
DB 48 FGALTAE 54

RESULT 50
S20528
hypothetical protein C - Synechococcus sp. (strain PCC 7942) plasmid pUH24
C:Species: Synechococcus sp.
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-May-2000
C:Accession: S20528
R:van der Plas, J.; Oosterhoff-Theerstra, R.; Borrijs, M.; Weisbeek, P.
Mol. Microbiol. 6, 653-664, 1992
A:Title: Identification of replication and stability functions in the complete nucleo
A:Reference number: S20525; MUID:9204021
A:Accession: S20528
A:Molecule type: DNA
A:Residues: 1-181 <PLA>
A:Cross-references: EMBL:S89470; NID:g247785; PIDN:AAB21872.1; PID:g247789
C:Genetics:
A:Genome: plasmid
A:Start codon: GTG
C:Superfamily: Synechococcus sp. plasmid pUH24 hypothetical protein C

Query Match 0.7%; Score 7; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 878 LEQLTAL 884
|||||
DB 66 LEQLTAL 72

RESULT 51
B64398
hypothetical protein MJ0786 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: B64398
R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: B64398
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-186 <BUU>
A:Cross-references: GB:U67523; GB:L77117; NID:g2826319; PIDN:AA898786.1; PID:g1591486
C:Genetics:
A:Map position: FOR710065-710625
A:Start codon: TTG

Query Match 0.7%; Score 7; DB 2; Length 186;

Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 591 IPDYLF 597
|||||
Db 140 IPDYLF 146

RESULT 52

S48994

hypothetical protein YHR213w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein YAR062w

C:Species: *Saccharomyces cerevisiae*

C>Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 29-Oct-1999

C:Accession: S48994; S53471

R:Macri, C.

Submitted to the EMBL Data Library, February 1994

A:Description: The sequence of *S. cerevisiae* cosmid 9177.

A:Reference number: S4671

A:Accession: S48994

A:Molecule type: DNA

A:Residues: 1-198 <MAC>

A:Cross-references: EMBL:U00029; MIPS:YHR213w; NID:g551322; PIDN:AAB69730.1; PID:g458918

A:Genetics: CHR8

R:Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac

submitted to the EMBL Data Library, February 1994

A:Description: Sequencing of chromosome I of *Saccharomyces cerevisiae*: analysis of the 5

A:Reference number: S53458

A:Accession: S53471

A:Molecule type: DNA

A:Residues: 1-198 <BUS>

A:Cross-references: EMBL:L28920; MIPS:YAR062w; NID:g1616966; PIDN:AAC09503.1; PID:g45615

A:Genetics: CHR1

C:Genetics: <CHR8>

A:Map position: 8R

C:Genetics: <CHR1>

A:Map position: 1R

Query Match

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 IVTTTTE 282
|||||
Db 136 IVTTTTE 142

RESULT 53

ABC transporter (ATP-binding protein) BH0814 [imported] - *Bacillus halodurans* (strain C-

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000

C:Accession: F83751

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20263314

A:Accession: F83751

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-210 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04533.1; GSPDB:GN00

C:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0814

Query Match

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180

Db 10 GKSTLLQ 16
|||||

RESULT 54

G81335

probable ABC transporter ATP-binding protein Cj1277c [imported] - *Campylobacter jejuni*

C:Species: *Campylobacter jejuni*

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 08-Sep-2000

C:Accession: G81335

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chil

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals

A:Reference number: A81250; MUID:20150912

A:Accession: G81335

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-221 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CA073531.1; PID:g696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj1277c

C:Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homolo

Query Match

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 236 LKLROR 242
|||||
Db 75 LKLROR 81

RESULT 55

F71309

probable phosphoglycolate phosphatase (gph-2) - *Syphilis spirochete*

C:Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)

C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999

C:Accession: F71309

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

erson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M

Science 281, 375-388, 1998

A:Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.

A:Reference number: A71250; MUID:98332770

A:Accession: F71309

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-222 <COL>

A:Cross-references: GB:AE001231; GB:AE000520; NID:g3322846; PIDN:AAC26558.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0554

C:Superfamily: Alkaligenes eutrophus phosphoglycolate phosphatase

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 SRALTOR 14
|||||
Db 52 SRALTOR 58

RESULT 56

B69670

choline ABC transporter (membrane protein) opubD - *Bacillus subtilis*

N:Alternate names: hypothetical protein (spat 5' region)

C:Species: *Bacillus subtilis*

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C/Accession: B69670; B45740
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Ertian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hoshono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadata, Y.; Sato, T.; Scanlon,
A:Authors: Schleicher, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serd
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Dancho, A.
A:Title: The complete genome sequence of the Gram positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: B69670
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-226 <KUN>
A:Cross-references: GB:299121; GB:AL009126; NID:92635827; PIDN:CAB15375.1; PID:92635883
A:Experimental source: strain 168
R:Chung, Y.J.; Hansen, J.N.
J. Bacteriol. 174, 6699-6702, 1992
A:Title: Determination of the sequence of spae and identification of a promoter in the s
A:Reference number: A45740; MUID:93015727
A:Accession: B45740
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 25-29, 'V', '31-42', 'V', '44-49', 'Y', '51-88', 'I', '90-107', 'V', '109-188', 'T', '190-195', 'I', '1
A:Cross-references: GB:W9263; NID:9143562
C:Genetics:
A:Gene: OPUBD
C:Superfamily: glycine betaine/carnitine/choline ABC transporter

Query Match 0.7%; Score 7; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 568 KSAISOE 574
|||||||
Db 109 KSAISOE 115

RESULT 58

C72105
conserved hypothetical protein CP0536 [imported] - *Chlamydia pneumoniae* (strains
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C/Accession: C72105; B81566
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.
A:Reference number: A72000; MUID:99206606
A:Accession: C72105
A:Molecule type: DNA
A:Residues: 1-233 <ARN>
A:Cross-references: GB:AE001608; GB:AE001363; NID:94376487; PIDN:AAD18381.1; PID:9437
A:Experimental source: strain CWJ029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: B81566
A:Molecule type: DNA
A:Residues: 1-233 <REA>
A:Cross-references: GB:AE002213; GB:AE002161; NID:97189446; PIDN:AAF8359.1; PID:9718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: dsbG; CP0536

Query Match 0.7%; Score 7; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGLK 788
|||||||
Db 148 KLAEGLK 154

RESULT 59
D86519
disulfide bond chaperone [imported] - *Chlamydia pneumoniae* (strain J138)
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C/Accession: D86519
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of *Chlamydia pneumoniae* J138.
A:Reference number: AB6491; MUID:20330349
A:Accession: D86519
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-233 <STO>
A:Cross-references: GB:BA000008; NID:98978601; PIDN:BA08438.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: dsbG

Query Match 0.7%; Score 7; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 782 KLAEGLK 788
|||||||
Db 148 KLAEGLK 154

RESULT 60
G75494
Azic family protein - *Deinococcus radiodurans* (strain RJ)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C/Accession: G75494
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896
A:Accession: G75494
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-235 <MUID>
A:Cross-references: GB:AE001921; GB:AE000513; NID:96458330; PIDN:AMF10212.1; PID:9645833
A:Experimental source: strain R1
A:Genetics:
A:Gene: DR0633
A:Map position: 1
C:Superfamily: hypothetical protein b2682

Query Match 0.7%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 RLPGGLT 756
|||||||
DB 121 RLPGGLT 127

RESULT 61
trac protein - *Escherichia coli* plasmid pKM101
C:Species: *Escherichia coli*
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Oct-1999
A:Accession: I79288; S61375
R:Polman, R.F.; Genetti, H.D.; Winans, S.C.
Plasmid 31, 158-165, 1994
A:Title: Entry exclusion of the IncN plasmid pKM101 is mediated by a single hydrophilic
A:Reference number: I58936; MUID:94302136
A:Accession: I79288
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-237 <RES>
A:Cross-references: EMBL:U09868; NID:9558074; PIDN:AAA66453.1; PID:9498877
R:Polman, R.F.; Genetti, H.D.; Winans, S.C.
Mol. Microbiol. 14, 655-668, 1994
A:Title: Common ancestry between IncN conjugal transfer genes and macromolecular export
A:Reference number: S61366; MUID:95198540
A:Accession: S61375
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-237 <POH>
A:Cross-references: EMBL:U09868; NID:9558074; PIDN:AAA66453.1; PID:9498877
A:Genetics:
A:Gene: trac
A:Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 420 VLLTGL 426
|||||||
DB 8 VLLTGL 14

RESULT 62
A64553
Probable holoxytochrome-c synthase (EC 4.4.1.17) HP0265 - *Helicobacter pylori* (strain 26
C:Species: *Helicobacter pylori*
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
A:Accession: A64553
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenna
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L.

Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:97394467
A:Accession: A64553
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-240 <TOM>
A:Cross-references: GB:AE000545; GB:AE000511; NID:92313349; PIDN:AMD07331.1; PID:9231
C:Keywords: carbon-sulfur lyase

Query Match 0.7%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 LFLKSLK 74
|||||||
DB 199 LFLKSLK 205

RESULT 63
E70959
hypothetical protein RV0207C - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
A:Accession: E70959
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: E70959
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-242 <COL>
A:Cross-references: GB:292669; GB:AL123456; NID:93242271; PIDN:CAB07002.1; PID:e30468
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV0207C

Query Match 0.7%; Score 7; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 455 LSSLTSS 461
|||||||
DB 234 LSSLTSS 240

RESULT 64
G84061
Ferriochrome ABC transporter (ATP-binding protein) BH3295 (imported) - *Bacillus halodur*
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
A:Accession: G84061
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
A:Reference number: A83650; MUID:20263314
A:Accession: G84061
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-256 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:910175792; PIDN:BA07014.1; GSPDB:G
A:Experimental source: strain C-125
A:Genetics:
A:Gene: BH3295

Query Match 0.7%; Score 7; DB 2; Length 256;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLQ 180
 |||||||
 Db 39 GKSTLQ 45

RESULT 65

T18151
 hypothetical protein A649R - Chlorella virus PBCV-1
 C:Species: Chlorella virus PBCV-1
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
 C:Accession: T18151
 R:Graves, M.V.; Van Etten, J.L.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: Z18806
 A:Accession: T18151
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1259 <GRA>
 A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96972.1
 A:Experimental source: specific host Chlorella strain NC64A
 C:Genetics:
 A:Note: A649R
 C:Superfamily: Chlorella virus PBCV-1 hypothetical protein A450R

Query Match 0.7%; Score 7; DB 2; Length 259;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 NLMKTP 339
 |||||||
 Db 159 NLMKTP 165

RESULT 66

T36431
 probable ABC-type transport system ATP-binding protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
 C:Accession: T36431
 R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A:Reference number: Z21598
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Accession: T36431
 A:Molecule type: DNA
 A:Residues: 1264 <SEE>
 A:Cross-references: EMBL:AL096837; PIDN:CAP48895.1; GSPDB:GNO0070; SCOEDB:SCF43A.08
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCF43A.08
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 0.7%; Score 7; DB 2; Length 264;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLQ 180
 |||||||
 Db 63 GKSTLQ 69

RESULT 67

A69102
 trehalase-6-phosphate phosphatase related protein - Methanobacterium thermoautotrophicum
 C:Species: Methanobacterium thermoautotrophicum
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: A69102

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
 Qiu, D.; Spadatore, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
 Kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
 A:Reference number: A69000; MUID:98037514
 A:Accession: A69102
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1264 <MTH>
 A:Cross-references: GB:AE000931; GB:AE000666; NID:g2622885; PIDN:AAB86226.1; PID:g262
 C:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1760

Query Match 0.7%; Score 7; DB 2; Length 264;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 592 PDYLFDF 598
 |||||||
 Db 2 PDYLFDF 8

RESULT 68

E72642
 hypothetical protein APE0573 - Aeropyrum pernix (strain KI)
 C:Species: Aeropyrum pernix
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: E72642
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeya, Y.; Jin-uo, K.; Ta
 wa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
 A:Reference number: A72450; MUID:99310339
 A:Accession: E72642
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1264 <KAW>
 A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79541.1; PID:d1043327; PID:g
 A:Experimental source: strain KI
 C:Genetics:
 A:Gene: APE0573

Query Match 0.7%; Score 7; DB 2; Length 264;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1012 LSVITGA 1018
 |||||||
 Db 61 LSVITGA 67

RESULT 69

S46534
 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) Rieske iron-sulfur protein - potato
 N:Alternate names: Rieske iron-sulfur protein
 C:Species: Solanum tuberosum (potato)
 C>Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 20-Apr-2000
 C:Accession: S46534; S45037
 R:Emmermann, M.; Clericus, M.; Braun, H.P.; Mozo, T.; Heins, L.; Kruff, V.; Schmitz,
 Plant Mol. Biol. 25, 271-281, 1994
 A:Title: Molecular features, processing and import of the Rieske iron-sulfur protein
 A:Reference number: S46534; MUID:94289650
 A:Accession: S46534
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1265 <EMM>
 A:Cross-references: EMBL:X79332
 A:Note: the authors translated the codon CCM for residue 215 as Thr and GAA for resid
 R:Emmermann, M.; Clericus, M.; Braun, H.P.; Mozo, T.; Heins, L.; Kruff, V.; Schmitz,

submitted to the EMBL Data Library, May 1994
 A:Description: Molecular features, processing and import of the Rieske iron sulfur prote
 A:Reference number: S45037
 A:Accession: S45037
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-118, 'SV', 121-265 <EM2>
 A:Cross-references: EMBL:X79332; NID:g488847; PIDN:CAA55894.1; PID:g488848
 C:Superfamily: ubiquinol--cytochrome-c reductase iron-sulfur protein; Rieske [2Fe-2S] h
 C:Keywords: 2Fe-2S; electron transfer; membrane-associated complex; metalloprotein; mlt
 F:198-245/Domains: Rieske [2Fe-2S] homology <RSK>
 F:208,210,227,230/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #status F
 F:213-229/Disulfide bonds: #status predicted
 F:230/Active site: His #status predicted

Query Match 0.7%; Score 7; DB 2; Length 265;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 451 AGRRLS 457
 |||||
 DB 5 AGRRLS 11

RESULT 70
 T29609
 hypothetical protein C54H2.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T29609
 R:Fulton, L.; Gattung, S.
 submitted to the EMBL Data library, May 1996
 A:Description: The sequence of C. elegans cosmid C54H2.
 A:Reference number: Z20651
 A:Accession: T29609
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-266 <FNU>
 A:Cross-references: EMBL:U58728; PIDN:AAB00589.1; GSPDB:GN00028; CESP:C54H2.4
 A:Experimental source: strain Bristol N2; clone C54H2
 C:Genetics:
 A:Gene: CESP:C54H2.4
 A:Map position: X
 A:Introns: 47/1; 76/3; 103/1; 187/3; 205/1; 238/1

Query Match 0.7%; Score 7; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 832 LSANAVK 838
 |||||
 DB 254 LSANAVK 260

RESULT 71
 A72401
 ABC transporter, ATP-binding protein - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: A72401
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; MUID:99287316
 A:Accession: A72401
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-266 <ARN>
 A:Cross-references: GB:AE001707; GB:AE000512; NID:g4980720; PIDN:AAD35314.1; PID:g498072

A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM0222
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 0.7%; Score 7; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLLQ 180
 |||||
 DB 45 GKSTLLQ 51

RESULT 72
 A5511
 2-hydroxypenta-2,4-dienoate hydratase - Pseudomonas sp. (strain KKS102)
 C:Species: Pseudomonas sp.
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 20-Jun-2000
 C:Accession: A5511
 R:Kikuchi, Y.; Yasukochi, Y.; Nagata, Y.; Fukuda, M.; Takagi, M.
 J. Bacteriol. 176, 4269-4276, 1994
 A:Title: Nucleotide sequence and functional analysis of the meta-cleavage pathway inv
 A:Reference number: A5511; MUID:94292455
 A:Accession: A5511
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-268 <KIK>
 A:Cross-references: GB:D16407; NID:g303760; PIDN:BA03891.1; PID:g425212
 C:Genetics:
 A:Gene: bphe
 C:Superfamily: 2-hydroxypenta-2,4-dienoate hydratase

Query Match 0.7%; Score 7; DB 2; Length 268;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 450 TAGRRLS 456
 |||||
 DB 57 TAGRRLS 63

RESULT 73
 S72382
 hypothetical protein 14 - Enterococcus faecalis plasmid PAD1
 C:Species: Enterococcus faecalis
 C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 15-Oct-1999
 C:Accession: S72382
 R:Hirt, H.; Wirth, R.; Muscholl, A.
 Mol. Gen. Genet. 252, 640-647, 1996
 A:Title: Comparative analysis of 18 sex pheromone plasmids from Enterococcus faecalis
 A:Reference number: S72375; MUID:97074879
 A:Accession: S72382
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-271 <HIR>
 A:Cross-references: EMBL:X96977; NID:g1279406; PIDN:CAA65673.1; PID:e236575; PID:g127
 A:Experimental source: strain OG1X
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996
 C:Genetics:
 A:Genome: plasmid PAD1

Query Match 0.7%; Score 7; DB 2; Length 271;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 379 KGVAASD 385
 |||||
 DB 44 KGVAASD 50

RESULT 74

A41607

ubiquinol--cytochrome-c reductase (EC 1.10.2.2) Rieske iron-sulfur protein precursor - N

N:Alternate names: Rieske iron-sulfur protein

C:Species: Zea mays (maize)

C:Date: 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change 20-Apr-2000

C:Accession: A41607

R:Huang, J.; Struck, F.; Matzinger, D.F.; Levings II, C.S.

Proc. Natl. Acad. Sci. U.S.A. 88, 10716-10720, 1991

A:Title: Functional analysis in yeast of cDNA coding for the mitochondrial Rieske iron-s

A:Reference number: A41607; MUID:92073358

A:Accession: A41607

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-273 <HUA>

A:Cross-references: GB:M77224; NID:9168606; PIDN:AAA3507.1; PID:9168607

C:Superfamily: ubiquinol--cytochrome-c reductase iron-sulfur protein; Rieske [2Fe-2S] hc

C:Keywords: 2Fe-2S; electron transfer; membrane-associated complex; metalloprotein; mlt

F:206-253/Domain: Rieske [2Fe-2S] homology <RSK>

F:216-218,235,238/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #status F

F:221-237/Disulfide bonds: #status predicted

F:238/Active site: His #status predicted

Query Match

Best Local Similarity 0.7%; Score 7; DB 2; Length 273;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 451 AGRRLSS 457

DB 5 AGRRLSS 11

RESULT 75

B69742

ABC transporter (ATP-binding protein) homolog ybaE - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001

C:Accession: B69742

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Enlilich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

iech, J.; Harwood, C.R.; Hentut, A.; Hiltner, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbel, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033

A:Accession: B69742

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-276 <KUN>

A:Cross-references: GB:Z99104; GB:AL009126; NID:92632267; PIDN:CAB11922.1; PID:el182079;

A:Experimental source: strain 168

C:Genetics:

A:Gene: ybaE

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C:Keywords: ATP, nucleotide binding, P-loop

F:10-209/Domain: ATP-binding cassette homology <ABC>

F:27-34/Region: nucleotide-binding motif A (P-loop)

Query Match 0.7%; Score 7; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 GKSTLQ 180

DB 32 GKSTLQ 38

Search completed: March 25, 2002, 11:01:36
Job time: 146 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS 5.00
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,020
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-207
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4103
TELEFAX: (517) 347-4103
TELEX: No. 5539095e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1225
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Ulmus Americana
STRAIN: NPS 3-487
INDIVIDUAL ISOLATE: N/A
DEVELOPMENTAL STAGE: N/A
HAPLOTYPE: N/A
TISSUE TYPE: N/A
CELL TYPE: N/A
CELL LINE: N/A
ORGANELLE: N/A
IMMEDIATE SOURCE: N/A
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: chitinase encoding DNA
LOCATION:
IDENTIFICATION METHOD: sequencing
OTHER INFORMATION: DNA needed for chitinase
OTHER INFORMATION: in elm.
PUBLICATION INFORMATION:
US-08-286-020-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-286-020-1 ..
Align seg 1/1 to: US-08-286-020-1 from: 1 to: 1225

452 G1YArgArgLeuSerIleuLeu 459
|||||
1085 GCGCGTGTGTCTGTCTCTCTT 1108

seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:us-08-603-919-1
seq_documentation_block:
Sequence 1, Application US/08603919
Patent No. 5728382
GENERAL INFORMATION:
APPLICANT: Masomeh B. Sticklen and Ravindra K. Hajela
TITLE OF INVENTION: A Chitinase cDNA Clone From a
TITLE OF INVENTION: Disease Resistant American
TITLE OF INVENTION: Elm Tree
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS 5.00
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,919
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-207
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4103
TELEFAX: (517) 347-4103
TELEX: No. 5728382e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1225
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Ulmus Americana
STRAIN: NPS 3-487
INDIVIDUAL ISOLATE: N/A
DEVELOPMENTAL STAGE: N/A
HAPLOTYPE: N/A
TISSUE TYPE: N/A
CELL TYPE: N/A
CELL LINE: N/A
ORGANELLE: N/A
IMMEDIATE SOURCE: N/A
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: chitinase encoding DNA
LOCATION:
IDENTIFICATION METHOD: sequencing
OTHER INFORMATION: DNA needed for chitinase
OTHER INFORMATION: in elm.
PUBLICATION INFORMATION:
US-08-603-919-1

alignment_scores:
Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x US-08-603-919-1 ..

Align seg 1/1 to: US-08-603-919-1 from: 1 to: 1225

452 GlyArgLeuSerLeu 459
|||||
1085 GGGCGTGTTCGTCACCTCTT 1108

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-272-875-1

seq_documentation_block:

Sequence 1, Application US/08272875
Patent No. 5487996
GENERAL INFORMATION:
APPLICANT: Takeji SHIBATANI
APPLICANT: Saburo KOMATSUBARA
APPLICANT: Kenji OMORI
APPLICANT: HIROYUKI AKATSUKA
TITLE OF INVENTION: GENE CODING FOR ESTERASE AND NOVEL
TITLE OF INVENTION: MICROORGANISM CONTAINING SAID GENE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/272,875
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/980,516A
FILING DATE: 25-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 355440/1991
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 137502/1992
FILING DATE: 13-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 20-3175P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703 241 1300
TELEFAX: 703 241 2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1839 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL:
ANTI-SENSE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN: Serratia marcescens Str41
US-08-272-875-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x US-08-272-875-1 ..

Align seg 1/1 to: US-08-272-875-1 from: 1 to: 1839

450 ThrAlaGlyArgLeuSer 457
|||||
1094 ACAGCGGCGCGACGTTATCATCG 1117

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-272-875-2

seq_documentation_block:

Sequence 2, Application US/08272875
Patent No. 5487996
GENERAL INFORMATION:
APPLICANT: Takeji SHIBATANI
APPLICANT: Saburo KOMATSUBARA
APPLICANT: Kenji OMORI
APPLICANT: HIROYUKI AKATSUKA
TITLE OF INVENTION: GENE CODING FOR ESTERASE AND NOVEL
TITLE OF INVENTION: MICROORGANISM CONTAINING SAID GENE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/272,875
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/980,516A
FILING DATE: 25-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 355440/1991
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 137502/1992
FILING DATE: 13-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 20-3175P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703 241 1300
TELEFAX: 703 241 2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1839 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL:
ANTI-SENSE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN: Serratia marcescens M-1
US-08-272-875-2

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-272-875-2 ..

Align seg 1/1 to: US-08-272-875-2 from: 1 to: 1839

450 ThrAAGyAArgLeuSerSer 457

|||||

1094 ACAGGGCCGACGTTTATCATCG 1117

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-350-741-1

seq_documentation_block:

Sequence 1, Application US/08350741

Patent No. 5804194

GENERAL INFORMATION:

APPLICANT: DOUGAN G.,

APPLICANT: CHARLES I. G.,

APPLICANT: HORMAECHE C.E.,

APPLICANT: JOHNSON K.S.,

APPLICANT: CHATFIELD S.N.,

TITLE OF INVENTION: LIVE VACCINES

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON and VANDERHVE PC

STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/350,741

FILING DATE: 05-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/952,737

FILING DATE: 09-MAY-1994

APPLICATION NUMBER: US 07/952,737

FILING DATE: 30-NOV-1992

APPLICATION NUMBER: GB 9007194.5

FILING DATE: 30-MAR-1990

APPLICATION NUMBER: PCT/GB91/00484

FILING DATE: 28-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: WILSON, MARY J.

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 117-158

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1980 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 395..1822

US-08-350-741-1

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-350-741-1/rev ..

Align seg 1/1 to reverse of: US-08-350-741-1 from: 1 to: 1980

451 AAGyAArgLeuSerSer 458

|||||

1331 GCTGGCGTCGACTTTCATCGCTT 1308

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-875A-1

seq_documentation_block:

Sequence 1, Application US/08463875A

Patent No. 5980907

GENERAL INFORMATION:

APPLICANT: DOUGAN, Gordon

APPLICANT: CHARLES, Ian G.

APPLICANT: HORMAECHE, Carlos E.

APPLICANT: JOHNSON, Kevin S.

APPLICANT: CHATFIELD, Steven N.

TITLE OF INVENTION: LIVE VACCINES

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON and VANDERHVE PC

STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,875A

FILING DATE: 05-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/340,741

FILING DATE: 07-DEC-1994

APPLICATION NUMBER: US 07/952,737

FILING DATE: 30-NOV-1992

APPLICATION NUMBER: GB 9007194.5

FILING DATE: 30-MAR-1990

APPLICATION NUMBER: PCT/GB91/00484

FILING DATE: 28-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: WILSON, MARY J.

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 117-158

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1980 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 395..1822

US-08-463-875A-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-463-875A-1/rev

Align seg 1/1 to reverse of: US-08-463-875A-1 from: 1 to: 1980

451 AlaGlyArgArgLeuSerLeu 458
|||||
1331 GCTGGCGTCGATTCATCGCTT 1308

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-353-585-4

seq_documentation_block:

Sequence 4, Application US/09353585
Patent No. 6287865

GENERAL INFORMATION:

APPLICANT: Dixon, Mark S
Jones, David A

TITLE OF INVENTION: Plant pathogen resistance genes and uses

thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC

STREET: 8th Floor, 1100 No. 6287865th Glebe Road

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/353,585

FILING DATE: 15-Jul-1999

CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q

1/68

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/930,277

FILING DATE: 27-Oct-1997

APPLICATION NUMBER: PCT/GB96/00785

FILING DATE: 01-Apr-1996

APPLICATION NUMBER: GB 9506658.5

FILING DATE: 31-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ms Mary J Wilson

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 620-69

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3573 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Tomato

STRAIN: Cf2

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-353-585-4

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-353-585-4

Align seg 1/1 to: US-09-353-585-4 from: 1 to: 3573

758 SerLeuGlyAsnLeuLysAsnLeu 765
|||||
980 TCATTGGGAGATCTGAAAAACTTG 1003

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-485-588-1

seq_documentation_block:

Sequence 1, Application US/08485588
Patent No. 5688938

GENERAL INFORMATION:

APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert

APPLICANT: Forrest H. Fuller

TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon

STREET: First Interstate World Center

STREET: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: FASTSEQ

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485,588

FILING DATE: 7 June, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: including application

PRIOR APPLICATION DATA: described below: 9

APPLICATION NUMBER: 08/353,784

FILING DATE: 9 December, 1994

APPLICATION NUMBER: PCT/US/94/12117

FILING DATE: 21 October, 1994

APPLICATION NUMBER: U.S. 08/292,827

FILING DATE: 23 August, 1994

APPLICATION NUMBER: U.S. 08/141,246

FILING DATE: 22 October, 1993

APPLICATION NUMBER: U.S. 08/009,389

FILING DATE: 23 February, 1993

APPLICATION NUMBER: U.S. 08/017,127

FILING DATE: 12 February, 1993

APPLICATION NUMBER: U.S. 07/934,161

FILING DATE: 21 August, 1992

APPLICATION NUMBER: U.S. 07/834,044

FILING DATE: 11 February, 1992

APPLICATION NUMBER: U.S. 07/749,451

FILING DATE: 23 August, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Hebert, Sheldon O.

REGISTRATION NUMBER: 38,179

REFERENCE/DOCKET NUMBER: 213/005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-485-588-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-485-588-1/rev ..

Align seg 1/1 to reverse of: US-08-485-588-1 from: 1 to: 5275

331 LeuArgAsnLeuMetLysThrPro 338
|||||
2433 TTGCGGAACTGTGATGAAGACGCC 2410

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-565-1

seq_documentation_block:
Sequence 1, Application US/08484565
Patent No. 5763569
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: Including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993

APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Hebert, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-484-565-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-484-565-1/rev ..

Align seg 1/1 to reverse of: US-08-484-565-1 from: 1 to: 5275

331 LeuArgAsnLeuMetLysThrPro 338
|||||
2433 TTGCGGAACTGTGATGAAGACGCC 2410

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-480-751-1

seq_documentation_block:
Sequence 1, Application US/08480751
Patent No. 5858684
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,751
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-480-751-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-480-751-1/rev ..
Align seg 1/1 to reverse of: US-08-480-751-1 from: 1 to: 5275

331 LeuArgAsnLeuMetLysThrPro 338
|||||
2433 TTGCGGAACTTGATGAAGACGCC 2410

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-943-986-1

seq_documentation_block:
Sequence 1, Application US/08943986
Patent No. 5962314
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,986
FILING DATE: 03-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,565
FILING DATE: 7-June-1995
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-943-986-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-943-986-1/rev ..
Align seg 1/1 to reverse of: US-08-943-986-1 from: 1 to: 5275

331 LeuArgAsnLeuMetLysThrPro 338
|||||
2433 TTGCGGAACTTGATGAAGACGCC 2410

seq_name: /cgn2_6/prodata/2/ina/6a_COMB.seq:US-08-353-784-1
seq_documentation_block:
; Sequence 1, Application US/08353784
; Patent No. 6011068
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert,
APPLICANT: Bradford C. Van Wagenen, Manuel
APPLICANT: F. Balandrin, Forrest H. Fuller,
APPLICANT: Eric G. Delmar, and Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
City: Los Angeles
State: California
Country: USA
Zip: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,784
FILING DATE: 9 December, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 8
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Hebert, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 209/069
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-353-784-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x US-08-353-784-1/rev ..
Align seg 1/1 to reverse of: US-08-353-784-1 from: 1 to: 5275
331 LeuAlgaShnLeuMetIysThrpPro 338
|||||
2433 TTGCGGAACTGTGATGAGACGCC 2410
seq_name: /cgn2_6/prodata/2/ina/6a_COMB.seq:US-08-484-719B-1
seq_documentation_block:
; Sequence 1, Application US/08484719B
; Patent No. 6031003
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert,
APPLICANT: Bradford C. Van Wagenen,
APPLICANT: Manuel F. Balandrin,
APPLICANT: Forrest H. Fuller, Eric G.
APPLICANT: Delmar, Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
City: Los Angeles
State: California
Country: USA
Zip: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS Word
SOFTWARE: FASTSEQ for Windows Version 3.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,719B
FILING DATE: 7 June, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Douglas C. Murdock
REGISTRATION NUMBER: 37,549
REFERENCE/DOCKET NUMBER: 213/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600

```

: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 5275 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 515..3769
: US-08-484-719B-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-484-719B-1/rev ..
Align seq 1/1 to reverse of: US-08-484-719B-1 from: 1 to: 5275

331 LeuArgAsnLeuMetLysThrPro 338
|||||
2433 TTGCGGAACTTGATGAAACGCCCC 2410

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-353-585-1

seq_documentation_block:
: Sequence 1, Application US/09353585
: Patent No. 6287865
: GENERAL INFORMATION:
:   APPLICANT: Dixon, Mark S
:             Jones, David A
:             Jones, Jonathan DG
: TITLE OF INVENTION: Plant pathogen resistance genes and uses
:             thereof
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Nixon & Vanderhye PC
:   STREET: 8th Floor, 1100 No. 6287865th Glebe Road
:   CITY: Arlington
:   STATE: Virginia
:   COUNTRY: United States of America
:   ZIP: 22201-4714
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/353,585
:   FILING DATE: 15-Jul-1999
:   CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q
:             1/68
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 08/930,277
:   FILING DATE: 27-Oct-1997
:   APPLICATION NUMBER: PCT/GB96/00785
:   FILING DATE: 01-Apr-1996
:   APPLICATION NUMBER: GB 9506658.5
:   FILING DATE: 31-Mar-1995
: ATTORNEY/AGENT INFORMATION:
:   NAME: Ms Mary J Wilson
:   REGISTRATION NUMBER: 32,955
:   REFERENCE/DOCKET NUMBER: 620-69
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (703) 816-4000
:   TELEFAX: (703) 816-4100
: INFORMATION FOR SEQ ID NO: 1:

```

```

: SEQUENCE CHARACTERISTICS:
:   LENGTH: 6471 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
:   ORGANISM: Tomato
:   STRAIN: Cf2
: FEATURE:
:   NAME/KEY: mat_peptide
:   LOCATION: 1754..5012
: FEATURE:
:   NAME/KEY: sig_peptide
:   LOCATION: 1677..1753
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-353-585-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-353-585-1 ..
Align seq 1/1 to: US-09-353-585-1 from: 1 to: 6471

758 SerLeuGlyAsnLeuLysAsnLeu 765
|||||
2664 TCATTGGGGAATCTGAATAACTTG 2687

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-061-376-4

seq_documentation_block:
: Sequence 4, Application US/08061376
: Patent No. 6175000
: GENERAL INFORMATION:
:   APPLICANT: Evans, Glen A.
:   APPLICANT: Djabali, Malek
:   APPLICANT: Selleri, Lucia
:   APPLICANT: Parry, Pauline
: TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11O23
:             TRANSLLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
:   STREET: 444 South Flower Street, Suite 2000
:   CITY: Los Angeles
:   STATE: California
:   COUNTRY: USA
:   ZIP: 90071
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   OPERATING SYSTEM: IBM PC compatible
:   SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/061,376
:   FILING DATE: 13-MAY-1993
:   CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
:   NAME: Reiter, Stephen E.
:   REGISTRATION NUMBER: 31,192
:   REFERENCE/DOCKET NUMBER: P41 9387
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (619)546-4737
:   TELEFAX: (619)546-9392
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 11907 base pairs

```

TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-061-376-4

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-061-376-4 ..

Align seg 1/1 to: US-08-061-376-4 from: 1 to: 11907

452 GYARGARGLeuSerLeuLeu 459
|||||
234 GGCGCGCGCTCAGCAGCTCTC 257

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-146-1

seq_documentation_block:

Sequence 1, Application US/08459146
Patent No. 586405
GENERAL INFORMATION:
APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,146
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12752 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryphonectria
parasitica)
STRAIN: EP713
US-08-459-146-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-459-146-1 ..

Align seg 1/1 to: US-08-459-146-1 from: 1 to: 12752

526 ARGPROLeuTPARGIncluser 533
|||||
7511 CGTCCTTTGGACAGCAATTC 7534

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-459-065-1

seq_documentation_block:

Sequence 1, Application US/08459065
Patent No. 5882642
GENERAL INFORMATION:
APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12752 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryphonectria
parasitica)
STRAIN: EP713
US-08-459-065-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-459-065-1 ..

Align seg 1/1 to: US-08-459-065-1 from: 1 to: 12752

526 ArgProLeuTrpArgGlnGluSer 533
|||||
7511 CGTCTCTTTGAGACAGGAATCC 7534

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-320-559-1

```
seq_documentation_block:
; Sequence 1, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaanl, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ANTI-SENSE: NO
; US-08-320-559-1
```

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-320-559-1 ..

Align seg 1/1 to: US-08-320-559-1 from: 1 to: 14255

452 GlyArgArgLeuSerSerLeuLeu 459
|||||
48 GGGCCGCCGCTCAGACGCTCCTC 71

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-327-392-1

```
seq_documentation_block:
; Sequence 1, Application US/08327392
; Patent No. 5633136
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaanl, Eli
; TITLE OF INVENTION: ALL-1 Polynucleotides and Monoclonal
; TITLE OF INVENTION: Antibodies for Leukemia Detection and
; TITLE OF INVENTION: Treatment
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633136r1s
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/327,392
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1331
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ANTI-SENSE: NO
; US-08-327-392-1
```

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-327-392-1 ..

Align seg 1/1 to: US-08-327-392-1 from: 1 to: 14255

452 GlyArgArgLeuSerSerLeuLeu 459

|||||
48 GGCCGCGCCTGACGACCTCTC 71

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-306-691B-55

seq_documentation_block:

Sequence 55: Application US/08306691B

Patent No. 5734039

GENERAL INFORMATION:

APPLICANT: Calabretta, Bruno

APPLICANT: Skorski, Tomasz

TITLE OF INVENTION: ANTISENSE

TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES

NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seidel, Gonda, Lavoragna & Monaco, P.C.

STREET: Two Penn Center, Suite 1800

City: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 720 KB

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/306,691B

FILING DATE: September 15, 1994

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Monaco, Daniel A.

REGISTRATION NUMBER: 30,480

REFERENCE/DOCKET NUMBER: 8321-8

TELEPHONE: (215) 568-8383

TELEFAX: (215) 568-5549

TELEX: NO. 5734039e

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 14255 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-306-691B-55

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-306-691B-55

Align seg 1/1 to: US-08-306-691B-55 from: 1 to: 14255

452 G1YArgArgLeuSerLeuLeu 459

|||||

48 GGCCGCGCCTGACGACCTCTC 71

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-545-860D-1

seq_documentation_block:

Sequence 1, Application US/08545860D

Patent No. 6040140

GENERAL INFORMATION:

APPLICANT: Crocco, Carlo

APPLICANT: Canaanl, Eli

TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

NUMBER OF INVENTION: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

STREET: One Liberty Place, 46th floor

City: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/545,860D

FILING DATE: 07-MAR-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/04496

FILING DATE: 22-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/10930

FILING DATE: 09-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/327,392

FILING DATE: 19-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/320,559

FILING DATE: 11-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/062,443

FILING DATE: 14-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,094

FILING DATE: 30-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/888,839

FILING DATE: 27-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/805,093

FILING DATE: 11-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: Deluca Esq., Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1262

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 14255

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

ANTI-SENSE: No

US-08-545-860D-1

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-545-860D-1

Align seg 1/1 to: US-08-545-860D-1 from: 1 to: 14255

452 G1YArgArgLeuSerLeuLeu 459

|||||

48 GGCCGCCGCTCAGACCTCTC 71
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US94-04496-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the A11-1
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: T0U-1242
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ANTI-SENSE: NO
; PCT-US94-04496-1
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x PCT-US94-04496-1 ..
Align seg 1/1 to: PCT-US94-04496-1 from: 1 to: 14255
452 G|A|A|G|A|G|L|E|U|S|E|R|S|E|R|L|E|U| 459
|||||
48 GGCCGCCGCTCAGACCTCTC 71
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq: US-08-781-891-79
seq_documentation_block:
; Sequence 79, Application US/08781891
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Hui
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO

;; TITLE OF INVENTION: WERNER'S SYNDROME
;; NUMBER OF SEQUENCES: 209
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SEED and BERRY LLP
;; STREET: 6300 Columbia Center, 701 Fifth Avenue
;; CITY: Seattle
;; STATE: Washington
;; COUNTRY: USA
;; ZIP: 98104-7092
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/781,891
;; FILING DATE: 27-DEC-1996
;; CLASSIFICATION: 800
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 6090620tenburg Ph.D., Carol
;; REGISTRATION NUMBER: 39,317
;; REFERENCE/DOCKET NUMBER: 240052.419
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 79:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 87350 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-781-891-79
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x US-08-781-891-79 ..
Align seg 1/1 to: US-08-781-891-79 from: 1 to: 87350
199 P|H|E|V|A|P|H|E|P|H|E|U|A|G|L|E|U|S|E|R| 206
|||||
82225 TTTGTTTTTTTTTGGAGACTGACT 82248
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq: US-08-840-767-29
seq_documentation_block:
; Sequence 29, Application US/08840767B
; Patent No. 6255464
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Riggs, Gregory J.
; TITLE OF INVENTION: Thlaagalingam, Sam
; FILE REFERENCE: 01107.05548
; CURRENT APPLICATION NUMBER: US/08/840,767B
; CURRENT FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: 60/015,823
; EARLIER FILING DATE: 1996-04-18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-840-767-29

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-840-767-29/rev ..

Align seg 1/1 to reverse of: US-08-840-767-29 from: 1 to: 34

216 leuCYASpGlnleuasp 222

|||||
 33 CTCTGTGACCACTGTTGCAT 13

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-356-790-2

seq_documentation_block:

; Sequence 2, Application US/08356790
 ; Patent No. 5586622
 ; GENERAL INFORMATION:
 ; APPLICANT: Gurr, Sarah J.
 ; APPLICANT: McPherson, Michael J.
 ; APPLICANT: Atkinson, Howard J.
 ; APPLICANT: Bowles, Diana J.
 ; TITLE OF INVENTION: Plant Parasitic Nematode Control
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Reising, Ethlington, Barnard, Perry & Milton
 ; STREET: P.O. Box 4390
 ; CITY: Troy
 ; STATE: Michigan
 ; COUNTRY: USA
 ; ZIP: 48099
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/356,790
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/988,954
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kohn, Kenneth I.
 ; REGISTRATION NUMBER: 30,995
 ; REFERENCE/DOCKET NUMBER: P-338 (UDL)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 689-3500
 ; TELEFAX: (810) 689-4071
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 110 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-356-790-2

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-356-790-2/rev ..

Align seg 1/1 to reverse of: US-08-356-790-2 from: 1 to: 110

847 leuValylsleuSerIleu 853

|||||

51 TTGGTAAATGTCCATTCTT 31

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-07-791-931-2

seq_documentation_block:

; Sequence 2, Application US/07791931C
 ; Patent No. 6133507
 ; GENERAL INFORMATION:
 ; APPLICANT: Raikhel, Natasha V.
 ; TITLE OF INVENTION: Nettle Lectin cDNA
 ; FILE REFERENCE: MSU 4.1-114
 ; CURRENT APPLICATION NUMBER: US/07/791,931C
 ; CURRENT FILING DATE: 1991-11-12
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO: 2
 ; LENGTH: 219
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: DNA fragment
 ; OTHER INFORMATION: containing parts of the UDA gene used as a probe
 ; OTHER INFORMATION: for the UDA gene.
 ; US-07-791-931-2

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-07-791-931-2/rev ..

Align seg 1/1 to reverse of: US-07-791-931-2 from: 1 to: 219

685 SerSerAlaThrSerLeuArg 691

|||||
 33 AGTTCGCCACCTCCTTGAGA 13

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-053-131-154

seq_documentation_block:

; Sequence 154, Application US/08053131
 ; Patent No. 5661016
 ; GENERAL INFORMATION:
 ; APPLICANT: Lomberg, Nils
 ; APPLICANT: Kay, Robert M.
 ; TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for
 ; NUMBER OF INVENTION: Producing Heterologous Antibodies
 ; NUMBER OF SEQUENCES: 197
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Kourie and Crew
 ; STREET: One Market Plaza, Stewart Tower, Suite 200
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94105
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/053,131
 ; FILING DATE: 26-APR-1993
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/990,860
 ; FILING DATE: 16-DEC-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/810,279
 ; FILING DATE: 17-DEC-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 14643-9-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-053-131-154

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-053-131-154/rev ..

Align seg 1/1 to reverse of: US-08-053-131-154 from: 1 to: 246

394 GYASPLEUALALEUGLUGLY 400
|||||
159 GGTGACCTGGCCTTGGAAGCA 139

seq_name: /cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-096-762-154

seq_documentation_block:

Sequence 154, Application US/08096762
Patent No. 5814318
GENERAL INFORMATION:
APPLICANT: Lomberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 5814318-Human Animals for
TITLE OF INVENTION: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Stewart Tower, Suite 200
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,762
FILING DATE: 22-JUL-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 14643-9-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-096-762-154

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-096-762-154/rev ..

Align seg 1/1 to reverse of: US-08-096-762-154 from: 1 to: 246

394 GYASPLEUALALEUGLUGLY 400
|||||
159 GGTGACCTGGCCTTGGAAGCA 139

seq_name: /cgn2_6/ptodata/2/1na/6B.COMB.seq:US-09-042-353-69

seq_documentation_block:

Sequence 69, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lomberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
TITLE OF INVENTION: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-042-353-69

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-69/rev ..
Align seg 1/1 to reverse of: US-09-042-353-69 from: 1 to: 246

394 G1YAspleuAlaIeuglucly 400
|||||
159 GGTGACCTGCGCTTGGAGA 139

seq_documentation_block:
; Sequence 333, Application US/08758417A
; Patent No. 6300129
; GENERAL INFORMATION:
```

```

APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 333:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 333:
US-08-758-417A-333

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-333/rev ..
Align seg 1/1 to reverse of: US-08-758-417A-333 from: 1 to: 246

394 G1YAspleuAlaIeuglucly 400
|||||
159 GGTGACCTGCGCTTGGAGA 139
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-94
seq_documentation_block:
Sequence 94, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
TITLE OF INVENTION: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-94
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x US-09-042-353-94/rev ..
Align seg 1/1 to reverse of: US-09-042-353-94 from: 1 to: 282
394 G|YAspleuAlaleuglugly 400
|||||
159 GTGACCTGGCCTTGCAAGCA 139
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-111
seq_documentation_block:
Sequence 111, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-113

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-113/rev ..

Align seg 1/1 to reverse of: US-09-042-353-113 from: 1 to: 282

394 G1yAspleuAlaleuGlugly 400
|||||
159 GGTGACCTGGCTTGAAGCA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-417A-358

seq_documentation_block:
Sequence 358, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993

APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-0090300S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 358:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 358:
US-08-758-417A-358

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-358/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-358 from: 1 to: 282

394 G1yAspleuAlaleuGlugly 400
|||||
159 GGTGACCTGGCTTGAAGCA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-417A-375

seq_documentation_block:
Sequence 375, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739

;; FILING DATE: 03-DEC-1993
;; APPLICATION NUMBER: US 08/155,301
;; FILING DATE: 18-NOV-1993
;; APPLICATION NUMBER: US 08/096,762
;; FILING DATE: 22-JUL-1993
;; APPLICATION NUMBER: US 08/053,131
;; FILING DATE: 26-APR-1993
;; APPLICATION NUMBER: US 07/990,860
;; FILING DATE: 16-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Serafini, Andrew T.
;; REGISTRATION NUMBER: 41,303
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 375:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 282 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; SEQUENCE DESCRIPTION: SEQ ID NO: 375:
US-08-758-417A-375

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-375/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-375 from: 1 to: 282

394 G1YAsPLeuAlaLeuGluGly 400
|||||
159 GGTGACCTGGCCTTGAAGGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-417A-377

seq_documentation_block:
; Sequence 377, Application US/08758417A
; Patent No. 6300129
; GENERAL INFORMATION:
; APPLICANT: Lomborg, Nils
; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
; Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 417
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,417A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995

;; APPLICATION NUMBER: US 08/352,322
;; FILING DATE: 07-DEC-1994
;; APPLICATION NUMBER: US 08/209,741
;; FILING DATE: 09-MAR-1994
;; APPLICATION NUMBER: US 08/165,699
;; FILING DATE: 10-DEC-1993
;; APPLICATION NUMBER: US 08/161,739
;; FILING DATE: 03-DEC-1993
;; APPLICATION NUMBER: US 08/155,301
;; FILING DATE: 18-NOV-1993
;; APPLICATION NUMBER: US 08/096,762
;; FILING DATE: 22-JUL-1993
;; APPLICATION NUMBER: US 08/053,131
;; FILING DATE: 26-APR-1993
;; APPLICATION NUMBER: US 07/990,860
;; FILING DATE: 16-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Serafini, Andrew T.
;; REGISTRATION NUMBER: 41,303
;; REFERENCE/DOCKET NUMBER: 014643-0090300S
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 377:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 282 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; SEQUENCE DESCRIPTION: SEQ ID NO: 377:
US-08-758-417A-377

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-377/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-377 from: 1 to: 282

394 G1YAsPLeuAlaLeuGluGly 400
|||||
159 GGTGACCTGGCCTTGAAGGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-78

seq_documentation_block:
; Sequence 78, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lomborg, Nils
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,353

FILED DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-78

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-09-042-353-78/rev ..
Align seq 1/1 to reverse of: US-09-042-353-78 from: 1 to: 284
394 GlyAspLeuAlaLeuGly 400
|||||
159 GGTGACCTGGCCTTGGAAGA 139

seq_name: /cgn2_6/pdata/2/lna/5B_COMB.seq:US-08-758-417A-342

seq_documentation_block:
Sequence 342, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lomborg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 342:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 342:
US-08-758-417A-342

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-758-417A-342/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-342 from: 1 to: 284

394 GYASPLAUAALaUgUgUly 400
 ||||||||||||||||||
 159 GGTGACCTGGCTTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-110

seq_documentation_block:

Sequence 110, Application US/09042353
 Patent No. 6255458
 GENERAL INFORMATION:
 APPLICANT: Lonberg, Nils
 TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
 NUMBER OF SEQUENCES: 421
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/042.353
 FILING DATE: 13-MAR-1998
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/810,279
 FILING DATE: 17-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/853,408
 FILING DATE: 18-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/904,068
 FILING DATE: 23-JUN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/990,860
 FILING DATE: 16-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/053,131
 FILING DATE: 26-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/096,762
 FILING DATE: 22-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/155,301
 FILING DATE: 18-NOV-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,739
 FILING DATE: 03-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/165,699
 FILING DATE: 10-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/209,741
 FILING DATE: 09-MAR-1994
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/352,322
 FILING DATE: 07-DEC-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/544,404
 FILING DATE: 10-OCT-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/728,463
 FILING DATE: 10-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US96/16433
 FILING DATE: 10-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/758,417
 FILING DATE: 02-DEC-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/21803
 FILING DATE: 01-DEC-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 014643-009040US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 110:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 285 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-09-042-353-110

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-09-042-353-110/rev ..

Align seg 1/1 to reverse of: US-09-042-353-110 from: 1 to: 285

394 GYASPLAUAALaUgUgUly 400
 ||||||||||||||||||
 159 GGTGACCTGGCTTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-112

seq_documentation_block:

Sequence 112, Application US/09042353
 Patent No. 6255458
 GENERAL INFORMATION:
 APPLICANT: Lonberg, Nils
 TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
 NUMBER OF SEQUENCES: 421
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/042.353

Tue Mar 26 11:36:12 2002

```

: FILING DATE: 13-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/810,279
: FILING DATE: 17-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/853,408
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/904,068
: FILING DATE: 23-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/990,860
: FILING DATE: 16-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/053,131
: FILING DATE: 26-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/096,762
: FILING DATE: 22-JUL-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/155,301
: FILING DATE: 18-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,739
: FILING DATE: 03-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/165,699
: FILING DATE: 10-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,741
: FILING DATE: 09-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/352,322
: FILING DATE: 07-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/544,404
: FILING DATE: 10-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/728,463
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US96/16433
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/758,417
: FILING DATE: 02-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US97/21803
: FILING DATE: 01-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 014643-009040US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO. 112:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 285 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-09-042-353-112

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:

```

```

US-09-697-089-2 x US-09-042-353-112/rev
Align seq 1/1 to reverse of: US-09-042-353-112 from: 1 to: 285
394 GtAspleuAlaLeuglucly 400
|||||
159 GGTGACCTGGCCTTGAGAGA 139
seq_name: /cgn2/6/plodata/2/ina/6B_COMB.seq:US-09-042-353-115
seq_documentation_block:
: Sequence 115, Application US/09042353
: Patent No. 6255458
: GENERAL INFORMATION:
: APPLICANT: Lonberg, Nils
: APPLICANT: Kay, Robert M.
: TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
: TITLE OF INVENTION: Producing Heterologous Antibodies
: NUMBER OF SEQUENCES: 421
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/042,353
: FILING DATE: 13-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/810,279
: FILING DATE: 17-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/853,408
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/904,068
: FILING DATE: 23-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/990,860
: FILING DATE: 16-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/053,131
: FILING DATE: 26-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/096,762
: FILING DATE: 22-JUL-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/155,301
: FILING DATE: 18-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,739
: FILING DATE: 03-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/165,699
: FILING DATE: 10-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,741
: FILING DATE: 09-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/352,322
: FILING DATE: 07-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/544,404
: FILING DATE: 10-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/728,463

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

159 GGTGACCTGGCCTTGGAAGGA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:us-09-042-353-118
seq_documentation_block:
; Sequence 118, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,353
; FILING DATE: 13-MAR-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US96/16433
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,417
; FILING DATE: 02-DEC-1996

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/21803
; FILING DATE: 01-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 014643-0090400S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 285 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-042-353-118

alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-118/rev ..
Align seg 1/1 to reverse of: US-09-042-353-118 from: 1 to: 285

394 GYASPLEUALAENGUGLY 400
|||||
159 GGTACCTGGCCTTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:us-08-758-417A-374
seq_documentation_block:
; Sequence 374, Application US/08758417A
; Patent No. 6300129
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
; NUMBER OF SEQUENCES: 417 Producing Heterologous Antibodies
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,417A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; APPLICATION NUMBER: US 08/161,739

FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 374:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 374:
US-08-758-417A-374

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-374/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-374 from: 1 to: 285

394 GYASPLEUALAUGLUGLY 400
|||||
159 GGTGACCTGGCCTTGAAGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-417A-376

seq_documentation_block:

Sequence 376, Application US/08758417A
Patent No. 6300129

GENERAL INFORMATION:

APPLICANT: Lonberg, Nils

Kay, Robert M.

TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies

NUMBER OF SEQUENCES: 417

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/758,417A

FILING DATE: 02-DEC-1996

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/728,463

FILING DATE: 10-OCT-1996

APPLICATION NUMBER: US 08/544,404

FILING DATE: 10-OCT-1995

APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 376:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 376:
US-08-758-417A-376

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-376/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-376 from: 1 to: 285

394 GYASPLEUALAUGLUGLY 400
|||||
159 GGTGACCTGGCCTTGAAGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-417A-379

seq_documentation_block:

Sequence 379, Application US/08758417A
Patent No. 6300129

GENERAL INFORMATION:

APPLICANT: Lonberg, Nils

Kay, Robert M.

TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies

NUMBER OF SEQUENCES: 417

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/758,417A

FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-0090300S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 379:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 379:
US-08-758-417A-379

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-379/rev ..
Align seg 1/1 to reverse of: US-08-758-417A-379 from: 1 to: 285

394 GYAAspleuAlaLeuGlucly 400
|||||
159 GGTGACCTGGCCTTGAAGCA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-758-417A-380

seq_documentation_block:
Sequence 380, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-0090300S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 380:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 380:
US-08-758-417A-380

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-380/rev ..
Align seg 1/1 to reverse of: US-08-758-417A-380 from: 1 to: 285

394 GYAAspleuAlaLeuGlucly 400
|||||
159 GGTGACCTGGCCTTGAAGCA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-758-417A-382

seq_documentation_block:
Sequence 382, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417

```

CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,659
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 382:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 382:
US-08-758-417A-382

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-382/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-382 from: 1 to: 285

394 G|YAspleuAlaIeugIuGly 400
|||||
159 G|GfACCTGgCCTTGGAAGCA 139

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-042-353-73
seq_documentation_block:
; Sequence 73, Application US/09042353

```

Patent No. 6255458

GENERAL INFORMATION:

APPLICANT: Lonberg, Nils

APPLICANT: Kay, Robert M.

TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for

TITLE OF INVENTION: Producing Heterologous Antibodies

NUMBER OF SEQUENCES: 421

CORRESPONDENCE ADDRESS:

ADDRESS: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/042,353

FILING DATE: 13-MAR-1998

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/810,279

FILING DATE: 17-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/853,408

FILING DATE: 18-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/904,068

FILING DATE: 23-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/990,860

FILING DATE: 16-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/053,131

FILING DATE: 26-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/096,762

FILING DATE: 22-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/155,301

FILING DATE: 18-NOV-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,739

FILING DATE: 03-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/165,699

FILING DATE: 10-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/209,741

FILING DATE: 09-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/352,322

FILING DATE: 07-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/544,404

FILING DATE: 10-OCT-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/728,463

FILING DATE: 10-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US96/16433

FILING DATE: 10-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/758,417

FILING DATE: 02-DEC-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/21803

FILING DATE: 01-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-73

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-73/rev ..

Align seg 1/1 to reverse of: US-09-042-353-73 from: 1 to: 287

394 GtAspleuAlaengLucly 400
|||||
159 GGTGACCTGGCCTTGAAGGA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-042-353-120

seq_documentation_block:
Sequence 120, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762

FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-120

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-120/rev ..

Align seg 1/1 to reverse of: US-09-042-353-120 from: 1 to: 287

394 GtAspleuAlaengLucly 400
|||||
159 GGTGACCTGGCCTTGAAGGA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-758-417A-337

seq_documentation_block:
Sequence 337, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417

```

CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 337:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 337:
US-08-758-417A-337

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-337/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-337 from: 1 to: 287

394 GlyaspLeuAlaLeuGluGly 400
|||||
159 GGTGACCTGGCTTGTGAAGA 139

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-08-758-417A-384
seq_documentation_block:
; Sequence 384, Application US/08758417A
```

```

Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 384:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 384:
US-08-758-417A-384

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-384/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-384 from: 1 to: 287

394 GlyaspLeuAlaLeuGluGly 400
```

159 GGTGACCTGGCCTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-109

seq_documentation_block:
; Sequence 109, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,353
; FILING DATE: 13-MAR-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US96/16433
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/758,417
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/21803
; FILING DATE: 01-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 014643-009040US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-042-353-109

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-109/rev ..

Align seg 1/1 to reverse of: US-09-042-353-109 from: 1 to: 288

394 GlyAspIeuAlaIeuGluGly 400
|||||
159 GGAGACCTGGCCTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-121

seq_documentation_block:
; Sequence 121, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,353
; FILING DATE: 13-MAR-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 121:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-121

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-121/rev ..
Align seg 1/1 to reverse of: US-09-042-353-121 from: 1 to: 288

394 GlysApLeuAlaLeuGluGly 400
|||||
159 GGTGACCTGGCTTGTGAAGA 139

seq_name: /cgn2_6/ploddata/2/1na/6B_COMB.seq:US-09-042-353-127
seq_documentation_block:
; Sequence 127, Application US/09042353

Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
TITLE OF INVENTION: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-127

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-09-042-353-127/rev ..

Align seg 1/1 to reverse of: US-09-042-353-127 from: 1 to: 288

394 GYAAspleuAlaLeuGlucly 400
|||||
159 GGAGACTGGCCTTGGAAGG 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-758-417A-373

seq_documentation_block:

Sequence 373, Application US/08758417A

Patent No. 6300129

GENERAL INFORMATION:

APPLICANT: Lonberg, Nils

KAY, Robert M.

TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for

Producing Heterologous Antibodies

NUMBER OF SEQUENCES: 417

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/758,417A

FILING DATE: 02-Dec-1996

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/728,463

FILING DATE: 10-OCT-1996

APPLICATION NUMBER: US 08/544,404

FILING DATE: 10-OCT-1995

APPLICATION NUMBER: US 08/352,322

FILING DATE: 07-DEC-1994

APPLICATION NUMBER: US 08/209,741

FILING DATE: 09-MAR-1994

APPLICATION NUMBER: US 08/165,699

FILING DATE: 10-DEC-1993

APPLICATION NUMBER: US 08/161,739

FILING DATE: 03-DEC-1993

APPLICATION NUMBER: US 08/155,301

FILING DATE: 18-NOV-1993

APPLICATION NUMBER: US 08/096,762

FILING DATE: 22-JUL-1993

APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 373:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 373:
US-08-758-417A-373

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x US-08-758-417A-373/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-373 from: 1 to: 288

394 GYAAspleuAlaLeuGlucly 400
|||||
159 GGAGACTGGCCTTGGAAGGA 139

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-758-417A-385

seq_documentation_block:

Sequence 385, Application US/08758417A

Patent No. 6300129

GENERAL INFORMATION:

APPLICANT: Lonberg, Nils

KAY, Robert M.

TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for

Producing Heterologous Antibodies

NUMBER OF SEQUENCES: 417

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/758,417A

FILING DATE: 02-Dec-1996

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/728,463

FILING DATE: 10-OCT-1996

APPLICATION NUMBER: US 08/544,404

FILING DATE: 10-OCT-1995

APPLICATION NUMBER: US 08/352,322

FILING DATE: 07-DEC-1994

APPLICATION NUMBER: US 08/209,741

FILING DATE: 09-MAR-1994

APPLICATION NUMBER: US 08/165,699

FILING DATE: 10-OCT-1998

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30

```


* CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/042,353
: FILING DATE: 13-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/810,279
: FILING DATE: 17-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/853,408
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/904,068
: FILING DATE: 23-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/990,860
: FILING DATE: 16-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/053,131
: FILING DATE: 26-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/096,762
: FILING DATE: 22-JUL-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/155,301
: FILING DATE: 18-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,739
: FILING DATE: 03-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/165,699
: FILING DATE: 10-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,741
: FILING DATE: 09-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/352,322
: FILING DATE: 07-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/544,404
: FILING DATE: 10-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/728,463
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US96/16433
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/758,417
: FILING DATE: 02-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US97/21803
: FILING DATE: 01-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 014643-009040US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 72:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 290 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-09-042-353-72

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-72/rev ..
Align seq 1/1 to reverse of: US-09-042-353-72 from: 1 to: 290
394 GYAspleuAlaleuglucly 400
|||||
159 GTGACCTGGCCTTGGAAGGA 139

seq_name: /cgn2/6/ptodata/2/lna/6B.COMB.seq:US-09-042-353-79

seq_documentation_block:
: Sequence 79, Application US/09042353
: Patent No. 6255458
: GENERAL INFORMATION:
: APPLICANT: Lonberg, Nils
: APPLICANT: Kay, Robert M.
: TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
: TITLE OF INVENTION: Producing Heterologous Antibodies
: NUMBER OF SEQUENCES: 421
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834

COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/042,353
: FILING DATE: 13-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/810,279
: FILING DATE: 17-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/853,408
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/904,068
: FILING DATE: 23-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/990,860
: FILING DATE: 16-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/053,131
: FILING DATE: 26-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/096,762
: FILING DATE: 22-JUL-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/155,301
: FILING DATE: 18-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,739
: FILING DATE: 03-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/165,699
: FILING DATE: 10-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,741
: FILING DATE: 09-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/352,322
: FILING DATE: 07-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/544,404
: FILING DATE: 10-OCT-1995

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/728,463
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US96/16433
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/758,417
: FILING DATE: 02-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US97/21803
: FILING DATE: 01-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 014643-009040US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 79:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 290 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-09-042-353-79

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-697-089-2 x US-09-042-353-79/rev  ..

Align seg 1/1 to reverse of: US-09-042-353-79  from: 1  to: 290

394 GYASPLEUALALEUGLGLY 400
|||||
159 GGTGACCTGGCCTTGGAAGA 139

seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-08-758-417A-336

seq_documentation_block:
: Sequence 336, Application US/08758417A
: Patent No. 6300129
: GENERAL INFORMATION:
: APPLICANT: Lonberg, Nils
: Kay, Robert M.
: TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
: Producing Heterologous Antibodies
: NUMBER OF SEQUENCES: 417
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/758,417A
: FILING DATE: 02-DEC-1996
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/728,463
: FILING DATE: 10-OCT-1996
```

```

: APPLICATION NUMBER: US 08/544,404
: FILING DATE: 10-OCT-1995
: APPLICATION NUMBER: US 08/352,322
: FILING DATE: 07-DEC-1994
: APPLICATION NUMBER: US 08/209,741
: FILING DATE: 09-MAR-1994
: APPLICATION NUMBER: US 08/165,699
: FILING DATE: 10-DEC-1993
: APPLICATION NUMBER: US 08/161,739
: FILING DATE: 03-DEC-1993
: APPLICATION NUMBER: US 08/155,301
: FILING DATE: 18-NOV-1993
: APPLICATION NUMBER: US 08/096,762
: FILING DATE: 22-JUL-1993
: APPLICATION NUMBER: US 08/053,131
: FILING DATE: 26-APR-1993
: APPLICATION NUMBER: US 07/990,860
: FILING DATE: 16-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Serafini, Andrew T.
: REGISTRATION NUMBER: 41,303
: REFERENCE/DOCKET NUMBER: 014643-009030US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 336:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 290 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 336:
  US-08-758-417A-336

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-697-089-2 x US-08-758-417A-336/rev  ..

Align seg 1/1 to reverse of: US-08-758-417A-336  from: 1  to: 290

394 GYASPLEUALALEUGLGLY 400
|||||
159 GGTGACCTGGCCTTGGAAGA 139

seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-08-758-417A-343

seq_documentation_block:
: Sequence 343, Application US/08758417A
: Patent No. 6300129
: GENERAL INFORMATION:
: APPLICANT: Lonberg, Nils
: Kay, Robert M.
: TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
: Producing Heterologous Antibodies
: NUMBER OF SEQUENCES: 417
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-00903005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 343:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 343:
US-08-758-417A-343

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-343/rev ..

Align seq 1/1 to reverse of: US-08-758-417A-343 from: 1 to: 290

394 GYAAspleuAlaLeugLucly 400
|||||
159 GGTGACCTGGCCTTGAAGGA 139

seq_name: /cgn2.6/ptodata/2/lna/6B-COMB.seq:US-09-042-353-89

seq_documentation_block:
Sequence 89, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
NUMBER OF INVENTIONS: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-00904005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 291 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

US-09-042-353-89

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-89/rev ..

Align seg 1/1 to reverse of: US-09-042-353-89 from: 1 to: 291

394 GYASPLEALALEUGLGLY 400
|||||
159 GGTGACCTGGCCTTGGAAGA 139

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-08-758-417A-353

seq_documentation_block:
Sequence 353, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lomborg, Nils
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 353:

SEQUENCE CHARACTERISTICS:
LENGTH: 291 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 353:
US-08-758-417A-353

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-353/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-353 from: 1 to: 291

394 GYASPLEALALEUGLGLY 400
|||||
159 GGTGACCTGGCCTTGGAAGA 139

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-09-042-353-71

seq_documentation_block:
Sequence 71, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lomborg, Nils
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 293 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-71

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-71/rev ..

Align seg 1/1 to reverse of: US-09-042-353-71 from: 1 to: 293

394 G1yAspleuAlaenglucly 400
|||||
159 GGTGACCTGGCTTGAAGCA 139

seq_name: /cgn2.6/ptodata/2/1na/6B-COMB.seq:US-09-042-353-75

seq_documentation_block:
Sequence 75, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
NUMBER OF INVENTIONS: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 293 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

US-09-042-353-75

Alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-75/rev ..

Align seg 1/1 to reverse of: US-09-042-353-75 from: 1 to: 293

394 GYASPLEUAlaLeuGlucly 400
|||||
159 GGTGACCTGCTTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-042-353-82

seq_documentation_block:
Sequence 82, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
NUMBER OF INVENTION: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741

FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 293 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-82

Alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-82/rev ..

Align seg 1/1 to reverse of: US-09-042-353-82 from: 1 to: 293

394 GYASPLEUAlaLeuGlucly 400
|||||
159 GGTGACCTGCTTGGAAGA 139

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-758-417A-335

seq_documentation_block:
Sequence 335, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 335:
SEQUENCE CHARACTERISTICS:
LENGTH: 293 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 335:
US-08-758-417A-335

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-335/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-335 from: 1 to: 293

394 G|A|S|P|l|e|u|a|l|e|u|g|l|u|g|y| 400
|||||
159 G|G|T|G|A|C|C|T|G|G|C|T|T|G|G|A|G|A| 139

seq_name: /cgn2_6/ptodata/2/ina/6B-COMB.seq:US-08-758-417A-339

seq_documentation_block:
Sequence 339, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 339:
SEQUENCE CHARACTERISTICS:
LENGTH: 293 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 339:
US-08-758-417A-339

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-08-758-417A-339/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-339 from: 1 to: 293

394 G|A|S|P|l|e|u|a|l|e|u|g|l|u|g|y| 400
|||||
159 G|G|T|G|A|C|C|T|G|G|C|T|T|G|G|A|G|A| 139

seq_name: /cgn2_6/ptodata/2/ina/6B-COMB.seq:US-08-758-417A-346

seq_documentation_block:
Sequence 346, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1 seq documentation block:
2 ; Sequence 122, Application US/09042353
3 ; Patent No. 6255458
4
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Lonberg, Nils
7 ; APPLICANT: Kay, Robert M.
8 ; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
9 ; TITLE OF INVENTION: Producing Heterologous Antibodies
10 ; NUMBER OF SEQUENCES: 421
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: Townsend and Townsend and Crew LLP
13 ; STREET: Two Embarcadero Center, Eighth Floor
14 ; CITY: San Francisco
15 ; STATE: California
16
17 ; COUNTRY: USA
18 ; ZIP: 94111-3834
19
20 ; COMPUTER READABLE FORM:
21 ; MEDIUM TYPE: Floppy disk
22 ; COMPUTER: IBM PC compatible
23 ; OPERATING SYSTEM: PC-DOS/MS-DOS
24 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
25
26 ; CURRENT APPLICATION DATA:
27 ; APPLICATION NUMBER: US/09/042,353
28 ; FILING DATE: 13-MAR-1998
29 ; CLASSIFICATION: 800
30
31 ; PRIORITY APPLICATION DATA:
32 ; APPLICATION NUMBER: US 07/810,279
33 ; FILING DATE: 17-DEC-1991
34
35 ; PRIORITY APPLICATION DATA:
36 ; APPLICATION NUMBER: US 07/853,408
37 ; FILING DATE: 18-MAR-1992
38
39 ; PRIORITY APPLICATION DATA:
40 ; APPLICATION NUMBER: US 07/904,068
41 ; FILING DATE: 23-JUN-1992
42
43 ; PRIORITY APPLICATION DATA:
44 ; APPLICATION NUMBER: US 07/990,860
45 ; FILING DATE: 16-DEC-1992
46
47 ; PRIORITY APPLICATION DATA:
48 ; APPLICATION NUMBER: US 08/053,131
49 ; FILING DATE: 26-APR-1993
50
51 ; PRIORITY APPLICATION DATA:
52 ; APPLICATION NUMBER: US 08/096,762
53 ; FILING DATE: 22-JUL-1993
54
55 ; PRIORITY APPLICATION DATA:
56 ; APPLICATION NUMBER: US 08/155,301
57 ; FILING DATE: 18-NOV-1993
58
59 ; PRIORITY APPLICATION DATA:
60 ; APPLICATION NUMBER: US 08/161,739
61 ; FILING DATE: 03-DEC-1993
62
63 ; PRIORITY APPLICATION DATA:
64 ; APPLICATION NUMBER: US 08/165,699
65 ; FILING DATE: 10-DEC-1993
66
67 ; PRIORITY APPLICATION DATA:
68 ; APPLICATION NUMBER: US 08/209,741
69 ; FILING DATE: 09-MAR-1994
70
71 ; PRIORITY APPLICATION DATA:
72 ; APPLICATION NUMBER: US 08/352,322
73 ; FILING DATE: 07-DEC-1994
74
75 ; PRIORITY APPLICATION DATA:
76 ; APPLICATION NUMBER: US 08/544,404
77 ; FILING DATE: 10-OCT-1995
78
79 ; PRIORITY APPLICATION DATA:
80 ; APPLICATION NUMBER: US 08/728,463
81 ; FILING DATE: 10-OCT-1996
82
83 ; PRIORITY APPLICATION DATA:
84 ; APPLICATION NUMBER: WO PCT/US96/16433
85 ; FILING DATE: 10-OCT-1996
86
87 ; PRIORITY APPLICATION DATA:
88 ; APPLICATION NUMBER: US 08/758,417
89 ; FILING DATE: 02-DEC-1996
90
91 ; PRIORITY APPLICATION DATA:
92 ; APPLICATION NUMBER: WO PCT/US97/21803
93 ; FILING DATE: 01-DEC-1997
94

```



```

: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 014643-009040US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 122:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 294 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-09-042-353-122

alignment_scores:
      Quality: 7.00      Length: 7
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-122/rev ..

Align seg 1/1 to reverse of: US-09-042-353-122 from: 1 to: 294

394 GtGACCTGGCTTGGAGAGA 139
|||||
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-353-126

seq_documentation_block:
: Sequence 126, Application US/09042353
: Patent No. 6255458
: GENERAL INFORMATION:
: APPLICANT: Londerg, Nils
: TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
: NUMBER OF SEQUENCES: 421
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/042.353
: FILING DATE: 13-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/810,279
: FILING DATE: 17-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/853,408
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/904,068
: FILING DATE: 23-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/990,860
: FILING DATE: 16-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/053,131
: FILING DATE: 26-APR-1993
```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/096,762
: FILING DATE: 22-JUL-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/155,301
: FILING DATE: 18-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,739
: FILING DATE: 03-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/165,699
: FILING DATE: 10-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,741
: FILING DATE: 09-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/352,322
: FILING DATE: 07-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/544,404
: FILING DATE: 10-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/728,463
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US96/16433
: FILING DATE: 10-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/758,417
: FILING DATE: 02-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US97/21803
: FILING DATE: 01-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 014643-009040US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 126:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 294 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-09-042-353-126

alignment_scores:
      Quality: 7.00      Length: 7
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x US-09-042-353-126/rev ..

Align seg 1/1 to reverse of: US-09-042-353-126 from: 1 to: 294

394 GtGACCTGGCTTGGAGAGA 139
|||||
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-417A-386

seq_documentation_block:
: Sequence 386, Application US/08758417A
: Patent No. 6300129
: GENERAL INFORMATION:
: APPLICANT: Londerg, Nils
: TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

seq_documentation_block:
: Sequence 390, Application US/08758417A
: Patent No. 6300129
: GENERAL INFORMATION:
:   APPLICANT: Lonberg, Nils
:   Kay, Robert M.
:   TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
:   Producing Heterologous Antibodies
:   NUMBER OF SEQUENCES: 417
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Townsend and Townsend and Crew LLP
:   STREET: Two Embarcadero Center, Eighth Floor
:   CITY: San Francisco
:   STATE: California
:   COUNTRY: USA
:   ZIP: 94111-3834
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/758,417A
:   FILING DATE: 02-Dec-1996
:   CLASSIFICATION: <Unknown>
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 08/728,463
:   FILING DATE: 10-OCT-1996
:   APPLICATION NUMBER: US 08/544,404
:   FILING DATE: 10-OCT-1995
:   APPLICATION NUMBER: US 08/352,322
:   FILING DATE: 07-DEC-1994
:   APPLICATION NUMBER: US 08/209,741
:   FILING DATE: 09-MAR-1994
:   APPLICATION NUMBER: US 08/165,659
:   FILING DATE: 10-DEC-1993
:   APPLICATION NUMBER: US 08/161,739
:   FILING DATE: 03-DEC-1993
:   APPLICATION NUMBER: US 08/155,301
:   FILING DATE: 18-NOV-1993
:   APPLICATION NUMBER: US 08/096,762
:   FILING DATE: 22-JUL-1993
:   APPLICATION NUMBER: US 08/053,131
:   FILING DATE: 26-APR-1993
:   APPLICATION NUMBER: US 07/990,860
:   FILING DATE: 16-DEC-1992
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Serafini, Andrew T.
:   REGISTRATION NUMBER: 41,303
:   REFERENCE/DOCKET NUMBER: 016643-009030US
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (415) 576-0200
:   TELEFAX: (415) 576-0300
:   INFORMATION FOR SEQ ID NO: 390:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 294 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)
:   SEQUENCE DESCRIPTION: SEQ ID NO: 390:
US-08-758-417A-390

```

```

alignment_scores:
    Quality: 7.00      length: 7
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-6977-089-2 x US-08-758-417A-390/rev ..

Align seg 1/1 to reverse of: US-08-758-417A-390 from: 1 to: 294

```

394 GlyAspleuAlaLeuglucly 400
 |||||
 159 GGTGACCTGGCCTTGAAGA 139

seq_name: /cgn2.6/ptodata/2/lna/6B.COMB.seq:US-09-042-353-91

seq_documentation_block:

Sequence 91, Application US/09042353
 Patent No. 6255458
 GENERAL INFORMATION:
 APPLICANT: Lonberg, Nils
 APPLICANT: Kay, Robert M.
 TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
 TITLE OF INVENTION: Producing Heterologous Antibodies
 NUMBER OF SEQUENCES: 421
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/042,353
 FILING DATE: 13-MAR-1998
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/810,279
 FILING DATE: 17-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/853,408
 FILING DATE: 18-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/904,068
 FILING DATE: 23-JUN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/990,860
 FILING DATE: 16-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/053,131
 FILING DATE: 26-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/096,762
 FILING DATE: 22-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/155,301
 FILING DATE: 18-NOV-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,739
 FILING DATE: 03-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/165,699
 FILING DATE: 10-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/209,741
 FILING DATE: 09-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/352,322
 FILING DATE: 07-DEC-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/544,404
 FILING DATE: 10-OCT-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/728,463
 FILING DATE: 10-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US96/16433

FILING DATE: 10-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/758,417
 FILING DATE: 02-DEC-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/21803
 FILING DATE: 01-DEC-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 014643-009040US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 91:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 297 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-09-042-353-91

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x US-09-042-353-91/rev ..

Align seg 1/1 to reverse of: US-09-042-353-91 from: 1 to: 297

394 GlyAspleuAlaLeuglucly 400
 |||||
 159 GGTGACCTGGCCTTGAAGA 139

THIS PAGE BLANK (USPTO)

OM of: US-09-697-089-2 to: EST: * out_format : pfs
Date: Mar 26, 2002 7:32 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+pat.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09697089/runat.25032002.105210.12369/app_query.fasta_1.1097
-DB=EST -OFMT=fastap -SUFFIX=oligo.rst -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human0.csi -LIST=100 -DOCALLIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=75 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEPSTLZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US09697089.@CGN1_1.4971 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-697-089-2
Query Length: 1024
Database: EST: *
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 1176.470000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb.gss:AQ0309404	-	183.00	3169.55	2.2e-167	552	AQ0309404 CTB1-E1-2528J13.TF CT
gb.gst1:AV119179	-	156.00	2659.42	3.4e-141	480	AV119179 AV719179 G1C Homo Sap1
gb.gst1:AI163394	-	114.00	1968.53	1.7e-100	364	AI163394 GX57501.1 NCI CGAP Pa
gb.gst1:AM337918	-	74.00	1272.96	7.6e-62	261	AM337918 he12h11.1 NCI CGAP Pa
gb.gst1:AV565315	+	60.00	1026.36	5.5e-48	371	AV565315 AV655315 G1C Homo Sap1
gb.gss:AQ030928	-	56.00	953.34	6.2e-44	553	AQ030928 RPTC11-93C9.TV RPTC-11
gb.gst1:BG10375	-	55.00	936.99	5.0e-43	476	BG10375 RST29913 Athysa RAGH
gb.gss:AQ112439	-	48.00	812.77	4.2e-36	630	AQ112439 CIT-HSP-2372C1.TF CIT-H
gb.gst1:AM118836	-	45.00	761.99	2.8e-33	509	AM118836 ha12h11.1 NCI CGAP K1
gb.gss:AQ030886	+	35.00	588.07	1.4e-23	470	AQ030886 HS.5378.B2.C12.SP6P.R1
gb.gst1:HB25984	-	33.00	558.78	9.9e-22	219	HB25984 RPTC11-78E1.TV RPTC-11
gb.gst1:HB25984	-	32.00	536.69	9.9e-21	412	HB25984 Y156G07.r1 Soares breast
gb.gst1:AI1023795	+	30.00	500.38	1.0e-18	499	AI1023795 ox08603.1 Soares.fela
gb.gss:A2360053	-	23.00	376.85	7.9e-12	602	A2360053 IM0103H1F Mouse.10kb
gb.gss:BS58691	-	18.00	289.20	6.0e-07	635	BS58691 CIT-HSP-2013L21.TF CIT-H
gb.gst1:BF003662	+	18.00	287.15	7.9e-07	840	BF003662 601862546E1 RIKEN full-16
gb.gst1:BB593094	+	11.00	173.87	1.61	251	BB593094 BB593094 RIKEN full-16
gb.gst1:BB593094	+	10.00	157.36	13.35	275	BB593094 BB593094 RIKEN full-16
gb.gst1:BB575989	+	10.00	155.75	16.41	221	BB575989 BB575989 RIKEN full-16
gb.gst1:AM401018	+	10.00	153.20	22.75	389	AM401018 LamdiGest148est.L.dig1
gb.gst1:AI122242	+	10.00	153.20	25.88	446	AI122242 GH04F06.x1 Soares.NFL
gb.gst1:BF0185330	+	10.00	147.67	46.28	826	BF0185330 601844132E1 NIH.MGC.54
gb.gst1:AM307272	+	9.00	150.27	33.15	54	AM307272 sfs4h07.y1 Gm-cl009.G1
gb.gst1:AM166132	-	9.00	141.14	106.94	187	AM166132 AM166132 Rice panicle
gb.gst1:AM526139	-	9.00	140.04	123.04	217	AM526139 18288 MAMC 1BOV Bos.G1
gb.gst1:AM58474	-	9.00	139.15	137.95	245	AM58474 SH09F07.y1 Gm-cl016.G1
gb.gst1:BB330264	+	9.00	137.81	163.83	294	BB330264 BB330264 RIKEN full-16
gb.gst1:R401040	+	9.00	137.78	164.35	295	R401040 yf93h02.s1 Soares infant
gb.gst1:AI169370	+	9.00	137.56	169.08	304	AI169370 wq97f05.x1 NCI CGAP GC
gb.gst1:BB24128	+	9.00	137.33	177.43	314	BB24128 WHE0076.E10.120ZS whea
gb.gst1:BB181383	+	9.00	137.19	174.31	320	BB181383 UI-R-DCO-bz1-d-08-0-UI
gb.gst1:BI280717	-	9.00	136.61	191.02	346	BI280717 UI-R-DCO-bz1-a-06-0-UI
gb.gst1:AI1360226	-	9.00	136.55	192.58	349	AI1360226 th78h09.x1 Soares.NHH
gb.gst1:AI1384131	-	9.00	136.22	200.89	365	AI1384131 te36h05.x1 Soares.NHH
gb.gst1:AA157463	+	9.00	135.88	209.70	382	AA157463 z050h05.x1 StrataGene
gb.gst1:AI1348317	+	9.00	135.81	211.77	386	AI1348317 q015h06.x1 NCI CGAP Lu
gb.gst1:BB62922	+	9.00	135.73	213.84	390	BB62922 UI-M-BG1-a1f-g-09-0-UI
gb.gst1:AA665992	-	9.00	135.69	214.87	392	AA665992 UI-M-BG1-a1f-g-09-0-UI

gb.gst1:BF940899	+	9.00	135.56	218.49	399	BF940899 hu61a10.x1 NCI CGAP
gb.gst1:R42493	+	9.00	135.55	219.00	400	R42493 y902g08.s1 Soares inf
gb.gst1:AI191648	+	9.00	135.47	221.07	404	AI191648 gd47b07.x1 Soares.f
gb.gss:AQ0889169	-	9.00	135.47	221.07	404	AQ0889169 HS_2161.B1.A01.T7C
gb.gst1:AM509287	+	9.00	134.98	235.49	432	AM509287 AL509287 Hordeum vu
gb.gst1:AM266497	+	9.00	134.98	235.49	432	AM266497 YAS1 gastric carcin
gb.gst1:BG159456	+	9.00	134.91	237.54	436	BG159456 OV2_8.D03.G1.A002.0
gb.gst1:AL510027	-	9.00	134.86	239.08	439	AL510027 AL510027 Hordeum vu
gb.gst1:AL513286	+	9.00	134.78	241.65	444	AL513286 y106c12.x1 NCI CGAP
gb.gst1:AM149096	+	9.00	134.71	243.70	448	AM149096 x10802.x1 NCI CGAP
gb.gss:A2598362	-	9.00	134.58	247.80	456	A2598362 IM0413G07F Mouse 10
gb.gst1:AI1743742	+	9.00	134.57	248.31	457	AI1743742 wq53c03.x1 Soares.N
gb.gst1:AL512027	+	9.00	134.55	248.82	458	AL512027 AL512027 Hordeum vu
gb.gst1:BF435974	+	9.00	134.53	249.34	459	BF435974 nab7h07.x1 Soares
gb.gst1:AM3032702	+	9.00	134.50	250.36	461	AM3032702 s469h02.x1 NCI CGAP
gb.gst1:AA032702	+	9.00	134.42	252.92	466	AA032702 m134f10.r1 Soares m
gb.gst1:BG055260	+	9.00	134.38	254.46	469	BG055260 nad5c01.x1 NCI CGAP
gb.gst1:BE517849	+	9.00	134.21	260.08	480	BE517849 WHE0803.B04.D07ZS W
gb.gst1:BI279982	-	9.00	134.14	262.12	484	BI279982 UI-R-DCO-cba-g-03-0
gb.gst1:AL509297	-	9.00	134.07	264.16	488	AL509297 AL509297 Hordeum vu
gb.gst1:AA156049	+	9.00	134.05	265.18	490	AA156049 z050h05.s1 StrataGene
gb.gst1:BI281408	+	9.00	134.04	265.69	491	BI281408 UI-R-DCO-bz1-f-10-0
gb.gst1:AI1697488	+	9.00	133.82	273.34	506	AI1697488 w14609.x1 NCI CGAP
gb.gst1:BI280561	-	9.00	133.82	273.34	506	BI280561 UI-R-DCO-bz1-f-10-0
gb.gst1:BI280626	-	9.00	133.82	276.90	513	BI280626 UI-R-DCO-bz1-f-09-0
gb.gst1:BI280541	-	9.00	133.59	281.48	522	BI280541 UI-R-DCO-bz1-f-01-0
gb.gst1:BI281386	-	9.00	133.57	284.02	523	BI281386 UI-R-DCO-bz1-d-11-0
gb.gst1:BI280985	-	9.00	133.52	284.02	527	BI280985 UI-R-DCO-bz1-d-11-0
gb.gst1:AM794880	+	9.00	133.46	286.05	531	AM794880 RC6-DM0015-280300-0
gb.gst1:AL5109343	-	9.00	133.45	286.56	532	AL5109343 AL5109343 Hordeum vu
gb.gst1:AL511550	-	9.00	133.41	288.09	535	AL511550 AL511550 Hordeum vu
gb.gst1:AV567320	+	9.00	132.85	309.36	577	AV567320 AV567320 Arabidops1
gb.gst1:BE875431	+	9.00	132.84	309.87	578	BE875431 601469152F1 NIH.MGC
gb.gss:A2367911	+	9.00	132.64	317.95	594	A2367911 IM0107D18 Mouse.10
gb.gst1:AM710297	+	9.00	132.60	319.46	597	AM710297 unpublishe
gb.gst1:BF272434	-	9.00	132.60	319.46	597	BF272434 CA_EB0014011f Goss
gb.gst1:BF484253	+	9.00	132.46	324.05	609	BF484253 WHE3221.A12.A23ZS W
gb.gst1:AL499880	+	9.00	132.43	325.51	609	AL499880 AL499880 Hordeum vu
gb.gst1:BF272289	+	9.00	132.40	328.03	614	BF272289 CA_EB0014H13f Goss
gb.gss:AQ486887	+	9.00	132.40	328.03	614	AQ486887 RPTC11-11-261P23.TJ R
gb.gst1:BF278089	+	9.00	132.21	336.08	630	BF278089 CA_EB0038B19f Goss
gb.gss:RH034799	+	9.00	132.16	338.10	634	RH034799 RPTC11-24-267U18.TJ R
gb.gst1:BE0501593	+	9.00	132.11	340.11	638	BE0501593 HVSMM0908M020f Hord
gb.gst1:BE0500347	+	9.00	132.07	342.12	642	BE0500347 P11.95.C09.g1.A002
gb.gst1:AL508326	+	9.00	132.02	342.62	643	AL508326 AL508326 Hordeum vu
gb.gst1:BE060564	+	9.00	131.98	346.15	650	BE060564 WHE0903.A01.B01ZS W
gb.gst1:BG415644	+	9.00	131.90	349.65	657	BG415644 HVSMM006011f Hord
gb.gst1:AL510571	-	9.00	131.88	350.65	659	AL510571 AL510571 Hordeum vu
gb.gst1:BI288044	-	9.00	131.84	352.16	662	BI288044 UI-R-DCO-cay-c-02-0
gb.gst1:AL511486	+	9.00	131.82	353.16	664	AL511486 AL511486 Hordeum vu
gb.gst1:BE704506	-	9.00	131.81	353.66	665	BE704506 Sc01_05h02.A Sc01.A

seq_name: gb.gss:AQ0309404
seq_documentation_block: 552 bp DNA
LOCUS AQ0309404 CTB1-E1-2528J13.TF C1B1-E1 Homo sapiens genomic clone 2528J13, 22-DEC-1998
DEFINITION CTB1-E1-2528J13.TF C1B1-E1 Homo sapiens genomic clone 2528J13, 22-DEC-1998
ACCESSION AQ0309404
VERSION AQ0309404.1 GI:4041438
KEYWORDS GSS.
SOURCE human.

```

ORGANISM      Homo sapiens
REFERENCE      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
                Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
                Venter,J.C.
TITLE          Use of a random human BAC End Sequence Database for Sequence-Ready
JOURNAL        Map Building
COMMENT        Unpublished (1998)
                Other-SSRs: CITBI-E1-2528J13.TR
                Contact: Shaying Zhao, William Nierman, Mark Adams
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850
                Tel: 301 838 0208
                Fax: 301 838 0208
                Email: hbe@tigr.org
                Clones are available from Research Genetics (info@resgen.com). BAC
                end search page:
                http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
                Seq primer: M13-21
                Class: BAC ends.
FEATURES       location/Qualifiers
                1..552
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="CITBI-E1"
                /sex="male"
                /cell_type="sperm"
                /note="Vector: pBelobAC11. Site_1: EcoRI; Site_2: EcoRI;
BASE COUNT    144 a 121 c 130 g 157 t
ORIGIN
alignment_scores:
    Quality: 183.00      Length: 183
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
    US-09-697-089-2 x AQ309404/rev ...
Align seg 1/1 to reverse of: AQ309404 from: 1 to: 552
477 MetValSerIleSerAspIleThrSerThyTyrSerSerLeuLeuArgTy 493
|||||
550 ATGGTTTCATTTCGGACATTACATCCACTATATAGCACCCCTCGCGTA 501
|||||
493 rThCysGlySerSerValGluAlaThrArgAlaValMetIysHisLeuA 510
|||||
500 CACCTGGGTCATCTGTGGAAGCACCGAGCTGTATGAAGCACCTCG 451
|||||
510 laAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAlaIysArg 526
|||||
450 CAGCAGGTATCAACACGGCTGCTCTCGAGCATTTCCATGCGCAAGAG 401
|||||
527 ProLeuTrpArgGlnGluSerLeuGlnSerValIysAsnThrThrGluG 543
|||||
400 CCTCTCGAGACAGGAATCTTTGCCAAAGTGTGAAAAACACCACTGACCA 351
|||||
543 nGluIleLeuIysAlaIleAsnIleAsnSerPheValGluCysGlyIleH 560
|||||
350 AGAATTTCTGAAGCCATTAACATCAATCTCTTGTAGAGTGGCATCC 301
|||||
560 tIsLeuTyrGlnGluSerThrSerIysSerAlaLeuSerGlnIlePheGlu 576
|||||
300 ATTTATATCAAGAGATGACATCAAAATCAGCCCTGAGCCCAAGAAATTGCA 251
|||||
577 AlaPheGlnGlyIysSerLeuTyrIleAsnSerGlyAsnIleProAs 593
|||||

```

```

250 GCTTCTTTCAAGGTAAAGCTTATATATCACTCAGGGAACATCCCGCA 201
593 pTyrLeuPheAspPhePheGlnHisLeuProAsnCysAlaSerAlaLeuA 610
|||||
200 TTACTATTGACTCTTCTTGAACATTTGCCCAATGTGCAAGTGTCTGCG 151
|||||
610 sPheIleIysLeuAspPheTyrGlyGlyAlaMetAlaSerTrpGluIys 626
|||||
150 ACTTCATTAAACTGACTTTATGGGGAGCATGCTTCATGCGAAAG 101
|||||
627 AlaAlaGluAspThrGlyGlyIleHisMetGlnGluAlaProGluThrTy 643
|||||
100 GCTGCAGAGACACAGAGTGGAAATCCATGGAAGAGGCCCGAGAAACCTA 51
|||||
643 rIleProSerArgAlaValSerLeuPheAsnTrpIysGlnIlePhe 659
|||||
50 CATCCCAAGACGAGGCTGTATCTTCTTCTTCACTGGAAGCAGGAATTC 2
seq_name: gb_est1:AV719179
seq_documentation_block:
LOCUS      AV719179      480 bp      mRNA      EST      16-OCT-2000
DEFINITION AV719179 GLC Homo sapiens cDNA clone GLCE0A10 5', mRNA sequence.
ACCESSION  AV719179
VERSION     AV719179.1 GI:10816331
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE  1 (bases 1 to 480)
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS    Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
            Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
            Y., Gu,Y., Chen,Z. and Han,Z.
            Homo sapiens cDNA GLC clones
            Unpublished (2000)
TITLE      Contact: Zeguang Han
JOURNAL    Chinese National Human Genome Center at Shanghai
COMMENT    351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES   location/Qualifiers
            1..480
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="GLCE0A10"
            /clone_lib="GLC"
            /tissue_type="corresponding non cancerous liver tissue"
            /dev_stage="Adult"
            /lab_host="SOLR"
            /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
BASE COUNT    138 a 120 c 85 g 137 t
ORIGIN
alignment_scores:
    Quality: 156.00      Length: 156
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
    US-09-697-089-2 x AV719179/rev ..
Align seg 1/1 to reverse of: AV719179 from: 1 to: 480
869 HisGluLeuIleAspArgMetAsnValLeuGlnGlnIleuThrAlaLeu 885
|||||
480 CATGAAGTGAAGAGATGAACGCTGTGAAGACGCTCACCGCATGTAT 431
|||||

```

```

885 tleuprotprpqlcysaspvalglnghlyserleuserleuleulysH 902
|||||
430 GCTGCCCTGGGGCTGTGACGTCCAGGCAAGCCGTGACGCTGTTGAAAC 381
|||||
902 lsleuglnghlyvalproghlnleuvalylsleughlyleulysasntrparg 918
|||||
380 ATTTGGAGAGAGTCCACCACTCGTCAAGCTTGGGTTGAAAAACGGAGA 331
|||||
919 leuthraspthrghlileargileleughlyalphepheghlylsasnpr 935
|||||
330 CTCACAGATACAGACATTAGATTATTTAGCTGATTTTGGAAAGAACCC 281
|||||
935 oleulysasnphghlnghlnleuasnleuaglylsnargylasertera 952
|||||
280 TCTGAAAAACTTCCAGCACTTGAATTTGGCGGAATCTGTGACAGACTG 231
|||||
952 spglytrpleualaphemetglyvalphegluasnleulysghlnleuval 968
|||||
230 ATGCATGGCTTGCCCTTCATGGGTGTATTTGAGAACTTAAAGCAATTACTG 181
|||||
969 phepheasphepserthrlysglupheleuproaspproalaleuvalar 985
|||||
180 TTTTGTGACTTTAGTAAAGAAATTTCTACCTGATCCAGCATTTAGTCAG 131
|||||
985 glysleuserghlnleuserlysleuthrphleughlnghlyalaeatgyl 1002
|||||
130 AAAACTTGGCCAGATGTATCCAACTTACTTTTTCGACAGAGCTAGGC 81
|||||
1002 euvalaglytrpghlnpheaspaspaspaspleuservalillethrglyala 1018
|||||
80 TTGTTGGGTGGCAATTTGATGATGATGATCTCAGTATTATACAGAGACT 31
|||||
1019 pheylsleuvalthrila 1024
|||||
30 TTAAACTAGTAACGTGCT 13
|||||
seq_name: gb_est1:AI263294
seq_documentation_block: 364 bp mRNA EST 03-FEB-1999
LOCUS AI263294
DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
mRNA sequence.
ACCESSION AI263294
VERSION AI263294.1 GI:3871497
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 364)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/MLNL at:
www.bio.lnl.gov/bdrrp/image/image.html
Insert length: 2146 Std Error: 0.00
Seq primer: -400P from Glibco
High quality sequence stop: 364.
Location/Qualifiers
1. 364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2005417"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"

```

```

/Note="Organ: pancreas; Vector: PCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT 117 a 84 c 55 g 108 t
ORIGIN
alignment_scores:
Quality: 114.00 Length: 114
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AI263294/rev ..
Align seg 1/1 to reverse of: AI263294 from: 1 to: 364
911 lysleughlyleulysasntrpargleuthraspthrghlileatgilele 927
|||||
364 AACCTTGGGTTGAAAAAACTGAGACTCACAGATACAGAGATTACAAATTT 315
|||||
927 ughlyalaphhepheghlylsasnproleulysasnphghlnleuasnL 944
|||||
314 AGGTGCATTTTGGAAAGAACCCCTGAAAAAAGCTTCCAGAGTTGAATT 265
|||||
944 euAlaglylsnargylaseraspghlytrpleualaphemetglyal 960
|||||
264 TGGCGGGAATCGGTGAGCAGTGTGATGATGCTGCTTCATGGCTGTA 215
|||||
961 phegluasnleulysghlnleuvalphepheasphepserthrlysgluPh 977
|||||
214 TTTGAGAACTTAAACAATTAGTGTGTTTTCACITTAAGTAAAGAAATT 165
|||||
977 eleuproaspproalaleuvalargylsleuserghlnleuserlysl 994
|||||
164 TCTACCTGATCCAGCATTAATGACAGAAACTTACCAAGCTTATACCAAGT 115
|||||
994 euthrphleughlnghlyalargyleuvalglytrpghlnpheaspasp 1010
|||||
114 TTAACCTTTCGTCAGAAAGCTAGGCTTGTGGGTGCAATTTGATGATGAT 65
|||||
1011 aspleuservalillethrglyalaphelyleuvalthrila 1024
|||||
64 GATCTCACTGTTATTACAGGTGCTTTTAACTAGTAACGTGCT 23
|||||
seq_name: gb_est1:AW337918
seq_documentation_block: 261 bp mRNA EST 31-JAN-2000
LOCUS AW337918
DEFINITION he12h11.x1 NCI_CGAP_CHL1 Homo sapiens cDNA clone IMAGE:2918853 3',
mRNA sequence.
ACCESSION AW337918
VERSION AW337918.1 GI:6834544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 261)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/MLNL at:

```

www.bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gjbco

High quality sequence step: 201.

FEATURES

Location/Qualifiers

1..261

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NCI-CGAP_CML1"

/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL

rearrangement positive, includes both chronic phase and

myeloid blast crisis"

/lab_host="DH10B"

/note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:

SalI; Site_2: NotI; Cloned unidirectionally. Primer:

Oligo dt. Library constructed by Life Technologies."

BASE COUNT

86 a 54 c 38 g 83 t

ORIGIN

alignment_scores:

Quality: 74.00 Length: 74
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AW337918/rev ..

Align seg 1/1 to reverse of: AW337918 from: 1 to: 261

951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAnsLeuLysGlnLe 967

260 AGGATGGATGGCTGCTCATCGTGTATTTAGAGATCTTAAGCAATT 211

967 uValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984

210 AGCTTTTGTGACTTGTACTTAAGAAATTTCTACGATCCAGCAATTAG 161

984 aAlrGlyLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000

160 TCAGAAACTTACCAAGTGTATCCAGTTAACTTTTCGCAAGAGCT 111

1001 ArgLeuValGlyTrpGlnPheAspAspAspLeuSerValIleThrG1 1017

110 AGGCTGTGGTGGCGCAATTGATGATGATGATGATGATGATGATGATG 61

1017 yAlaPheLysLeuValThrAla 1024

60 TCCTTTAACTAGTACTGCT 39

seq_name: gb_est1:AV656315

seq_documentation_block:

LOCUS AV656315 371 bp mRNA EST 07-SEP-2000
DEFINITION AV656315 G1C Homo sapiens cDNA clone G1CEQA10 3', mRNA sequence.
ACCESSION AV656315
VERSION AV656315.1 GI:9877329

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Qian, B., Wu, T., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,
Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Qu, J.,
Zeng, L., Xu, S., Gu, M., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu,
G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.
Homo sapiens cDNA clone
Unpublished (2000)

TITLE

JOURNAL

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

Location/Qualifiers

1..371

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="G1CEQA10"

/clone_lib="G1C"

/tissue_type="corresponding non cancerous liver tissue"

/dev_stage="Adult"

/lab_host="SOLR"

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

XhoI"

BASE COUNT

112 a 85 c 91 g 82 t 1 others

ORIGIN

alignment_scores:

Quality: 60.00 Length: 60
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AV656315 ..

Align seg 1/1 to: AV656315 from: 1 to: 371

679 TyrLeuGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIlely 695

163 TATCTGGGGAAATATTCACCTCTGCCACCAAGCCTCAGGCTGCAAAATAA 212

695 SARCYsAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysL 712

213 GAGATGCTGCTGGTGGTGGAGCCCTCAGTTGGTCTCAGCAGCTGTA 262

712 yAsnIleTySerLeuMetValGluAlaSerProLeuThrIleGluAsp 728

263 AGAACATTTATCTCTCATGTGGAAGCCAGTCCCTCACCACATAGAAGAT 312

729 GluArgHisIleThrSerValThrAsnLeu 738

313 GAGAGGCACATCACATCTGTACCAACCTG 342

seq_name: gb_gss:AO320928

seq_documentation_block:

LOCUS AO320928 553 bp DNA GSS 06-MAY-1999
DEFINITION RPC111-93C9.TV RPC111 Homo sapiens genomic clone RPC111-93C9, DNA
sequence.
ACCESSION AO320928
VERSION AO320928.1 GI:4053662

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., White, C., de Jong, P., and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Other GSSs: RPC111-93C9.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850

TITLE

JOURNAL

COMMENT

Clones are derived from the human BAC library RPC111. For BAC
library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://resgen.com>). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..553

/organism="Homo sapiens"
/db_xref="GDB:7535384"
/db_xref="taxon:9606"
/clone_lib="RPCI-11-93C9"
/clone_1lb="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"

BASE COUNT 170 a 107 c 114 g 162 t
ORIGIN

alignment_scores:
Quality: 56.00 Length: 56
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AQ320928/rev ..

Align seg 1/1 to reverse of: AQ320928 from: 1 to: 553

```

873 AsparqmetasValleuGlueThralaleuMetleuProtpg1 889
|||||
256 GACAGGATGAACGTGCTAGAACACTCACCGCATGCTGCCCTGGGG 207
|||||
889 yCyaspvalcInglSerleuSerleuLeuLysHisleuGluuv 906
|||||
206 CTGAGAGTGCAGCAGCCTGACGACGCTGTGAAACATTGGAGAGG 157
|||||
906 aLpProGInleuValleuLysleuLysAsnTPargleuThraspThr 922
|||||
156 TCCCACTACCTGTCACCTGGCTTGAACCTGAGACTCACAGATACA 107
|||||
923 GlutleArgileuGly 928
|||||
106 GAGATTAGATTAGCT 89

```

seq_name: gb_est2:BG210375

seq_documentation_block: 476 bp mRNA EST 21-APR-2001
LOCUS BG210375
DEFINITION RST39813 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 476)
AUTHORS Harrington,J.D., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E., Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J., and Ducar,M.

TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596

Email: scain@atersys.com
High quality sequence stop: 360.
Location/Qualifiers

FEATURES
source 1..476

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 120 a 99 c 107 g 149 t
ORIGIN

alignment_scores:
Quality: 55.00 Length: 55
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BG210375 ..

Align seg 1/1 to: BG210375 from: 1 to: 476

```

959 GlyValPheGluAsnLeuLysGlnLeuValPhePheAspPheSerThrly 975
|||||
244 GGTGTAATTGAGAACTTAACCAATTAGTGTTTGTGACTTACTACTAA 293
|||||
975 sGluPheLeuProAspProAlaLeuValArgLysLeuSerGlnValleus 992
|||||
294 AGAATTTCTACCTGATCCAGCATTAAGTCAAGAACTTAAGCAAGTGTAT 343
|||||
992 eLysLeuThPheLeuGlnGluAlaArgLeuValIGlyTPGlpPheAsp 1008
|||||
344 CCAAGTTAATTCTTCTCCAAAGACCTAGGCTTGTGGGTGCATTGAT 393
|||||
1009 AspAspAspLeuSer 1013
|||||
394 GATGATGATCTCACT 408

```

seq_name: gb_gss:AQ112439

seq_documentation_block: 630 bp DNA GSS 29-AUG-1998
LOCUS AQ112439
DEFINITION CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA sequence.
ACCESSION AQ112439
VERSION AQ112439.1 GI:3484599
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 630)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL unpublished (1998)
COMMENT Other-GSSs: CIT-HSP-2372C1.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org

Clones are available from Research Genetics (inforesgen.com). BAC

(prečišćena)
BACPAC Reson

or from Research Genetics (info@resgen.com). BAC end Web Server:

<http://www.htsc.washington.edu>

Plate: 954 row: F column: 24

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 470.

Location/Qualifiers

1. .470

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-954 Col-24 Row-F"

/clone_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT 142 a 101 c 107 g 117 t 3 others
ORIGIN

alignment_scores:

Quality: 35.00 Length: 35

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AQ624020 ..

Align seq 1/1 to: AQ624020 from: 1 to: 470

679 TyrluendlylslllepheserSerAlaThrSerleuargleuGlnIlely 695
|||||
65 TATCTGGGGAATAATTCAGCTGCGACAGCCTGCAAAATGAA 114
695 sargCyAlaGlyAlaGlySerleuSerleuValleuSerthrcysL 712
|||||
115 GAGATGTCGTGCTGTGGCTGGAAGCTCAGTTTGCTCTCAGCAGCTGTA 164

712 ysasn 713
|||||
165 AGAAC 169

seq_name: gb_gss:AQ283886

seq_documentation_block:

LOCUS AQ283886 219 bp DNA GSS 27-APR-1999

DEFINITION RPC111-78E13.TV RPCI-11 Homo sapiens genomic clone RPCI-11-78E13,

DNA sequence.

ACCESSION AQ283886

VERSION AQ283886.1 GI:3910204

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 219)

Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,

Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.

Use of human BAC End Sequences for Sequence-Ready Map Building

Unpublished (1998)

CONTACT: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .219

/organism="Homo sapiens"

/db_xref="GDB:7529676"

/db_xref="taxon:9606"

/clone="RPCI-11-78E13"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; RPC111 Human Male BAC Library"

BASE COUNT 69 a 45 c 46 g 59 t
ORIGIN

alignment_scores:

Quality: 33.00 Length: 33

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AQ283886 ..

Align seq 1/1 to: AQ283886 from: 1 to: 219

701 AlaGlySerleuSerleuValleuSerThrcyslysasnIleYrSerle 717
|||||
3 GCTGGAAGCCTTACTGTGTCCTCAGCAGCCTGTAAGAACATTTATCTCT 52
717 UmetValGluAlaSerProleuthrIleGluaspGluargHisIleThr 733
|||||
53 CATGCTGGAAGCCAGTCCTCCATGACATAGAAAGAGGACACATCACA 101

seq_name: gb_est2:H25984

seq_documentation_block:

LOCUS H25984 412 bp mRNA EST 10-JUL-1995

DEFINITION Y156g07.r1 Soares breast 3bDbst Homo sapiens cDNA clone

IMAGE:162300 5', mRNA sequence.

ACCESSION H25984

VERSION H25984.1 GI:895107

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 412)

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman

,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maita,M., Parsons,J.,

Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston

,R., Williamson,A., Mohlmann,P. and Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilison RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 942

High quality sequence stops: 347

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 942 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 347.

Location/Qualifiers

1. .412

FEATURES

source

```
/organism="Homo sapiens"
/db_xref="GDB:576544"
/db_xref="taxon:9606"
/clone="IMAGE:162300"
/clone_lib="Soares breast 3NBHst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and Eco RI sites of a
modified pT73 vector (Pharmacia). Library went through
one round of normalization to a Cot = 20. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      129 a      84 c      73 g      124 t      2 others
ORIGIN
```

```
alignment_scores:
    Quality:      32.00      Length:      32
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000
```

```
alignment_block:
US-09-697-089-2 x H25984/rev ..
```

```
Align seg 1/1 to reverse of: H25984 from: 1 to: 412
```

```
993 LysleuthrPhelenuIngluAlaArgLeuValGlyTrpGlnPheaspas 1009
|||||
338 AAGTTAACTTTCTGCAAGACGCTGTTGGTGGCAATTATGTA 289
```

```
1009 pAspAspleuSerValIleThrGlyAlaPheLysleuValIleAla 1024
|||||
```

```
288 TGATGATCTCAGTGTATTACAGTGCTTTTAAACTAGTACTGCT 243
```

```
seq_name: gb_est1:AI023795
```

```
seq_documentation_block:
LOCUS      AI023795      499 bp      mRNA      EST      28-AUG-1998
DEFINITION oX08d03.x1 Soares-fetal_liver_spleen_1NFLS-S1 Homo sapiens cDNA
clone IMAGE:1655717 3', mRNA sequence.
ACCESSION  AI023795
VERSION     AI023795.1 GI:3238839
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

```
REFERENCE  1 (bases 1 to 499)
AUTHORS   NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
UNPUBLISHED (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1179 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 439.
```

```
FEATURES
Source
Location/Qualifiers
1..499
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1655717"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
```

```
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer (5'
AACTGGAAGAATTAATTAAGAATCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and Eco RI sites of a
modified pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      144 a      109 c      98 g      148 t
ORIGIN
```

```
alignment_scores:
    Quality:      30.00      Length:      30
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000
```

```
alignment_block:
US-09-697-089-2 x AI023795 ..
```

```
Align seg 1/1 to: AI023795 from: 1 to: 499
```

```
89 LeupheHisGlnThrSerGluLysAspleuAspAspleuAlaGlnasple 105
|||||
409 CTTTTCATCAGACATCAGACAGAGCTTGACGATTTGGCTCAGAGATT 458
```

```
105 uLysAspleuTyRHISthPProSerPheLeuAsnPhetyr 118
|||||
```

```
459 AAAGGACTTGTCACATACCCCATCTTTCTGAACTTTAT 498
```

```
seq_name: gb_gss:A2360053
```

```
seq_documentation_block:
LOCUS      A2360053      602 bp      DNA      GSS      02-OCT-2000
DEFINITION IM0103H1F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0103H1 F, DNA sequence.
ACCESSION  A2360053
VERSION     A2360053.1 GI:10473753
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE  1 (bases 1 to 602)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly
,M., Rose,M., Rose,R., Stokes,R., Tingley,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
UNPUBLISHED (2000)
CONTACT: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 602.
```

```
FEATURES
Source
Location/Qualifiers
1..602
```

```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
```

```

/clone="UUGC1M0103H11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pMD29v; Purified genomic DNA from M. musculus C57Bl/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT      139 a      153 c      139 g      171 t
ORIGIN

alignment_scores:
    Quality:      23.00      Length:      23
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AZ360053/rev ..

Align seq 1/1 to reverse of: AZ360053 from: 1 to: 602

439 TTTTSPHPHPHHISLYSPPHGLGLTYRTHALAGIARGATG 455
|||||
416 TATTAATTTCTCATTAATCATTTGAGAGTACGCGAGGTGGAGACT 367
|||||
455 userSerLeuLeuThrSer 461
|||||
366 CACGACGTTGCTGACGCRCC 348

seq_name: gb_gss:B58691

seq_documentation_block:
LOCUS      B58691      635 bp      DNA      GSS      20-JUN-1998
DEFINITION CIT-HSP-2013L21.TF CIT-HSP Homo sapiens genomic clone 2013L21, DNA
sequence.
ACCESSION      B58691
VERSION      B58691.1
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 635)
AUTHORS      Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE      Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL      Unpublished (1997)
COMMENT      Other_GSSs: CIT-HSP-2013L21.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

```

```

Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

FEATURES
    source
        1..635
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="GDB:704263"
                /db_xref="taxon:9606"
                /clone="2013L21"
                /clone_lib="CIT-HSP"
                /sex="Male"
                /cell_type="Sperm"
                /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
                HindIII"

BASE COUNT      217 a      122 c      90 g      206 t
ORIGIN

alignment_scores:
    Quality:      18.00      Length:      18
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x B58691 ..

Align seq 1/1 to: B58691 from: 1 to: 635

911 LYSLEUGLYLEUYSASNTTPARGLEUThrAspPhrGluIleArgIle 927
|||||
2 AAGCTTGCGTTGAAAACCTGAGACTCAGACATACAGAGATTAGAAATTT 51
|||||
927 ugly 928
|||||
52 AGGT 55

seq_name: gb_est2:BF207840

seq_documentation_block:
LOCUS      BF207840      840 bp      mRNA      EST      06-NOV-2000
DEFINITION 601862546F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082074 5',
mRNA sequence.
ACCESSION      BF207840
VERSION      BF207840.1
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 840)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.lnl.gov
Plate: L1CM944 row: 1 column: 11
High quality sequence stop: 636.
Location/Qualifiers
    1..840
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4082074"
        /clone_lib="NIH_MGC_53"
        /tissue_type="carcinoma, cell line"
        /lab_host="DH10B (T1 phage-resistant)"

```


Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1. 221
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, adult male corpora
quadrigemina"
/sex="male"
/tissue_type="corpora quadrigemina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
transcribed by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 57 a 42 c 67 g 55 t
ORIGIN

alignment_scores:

Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BB575989 ..

Align seg 1/1 to: BB575989 from: 1 to: 221

169 glyluserglylserThrlenu 178
|||||
72 GGCGAATCCGGAAGGAAAGCACACTC 101

seq_name: gb_estl:BB575989

seq_documentation_block:

LOCUS BB575989 275 bp mRNA EST 30-NOV-2000
DEFINITION BB575989 RIKEN full-length enriched, 10 days neonate head Mus
musculus cDNA clone 5530600M20 5', mRNA sequence.
ACCESSION BB575989
VERSION BB575989.1 GI:11472533

KEYWORDS

EST.
house mouse.
mus musculus

REFERENCE

1 (bases 1 to 275)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodozama, Y.,
Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Kono
H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K.,
Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A.,
Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
T., Toya, T., Watahiki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,

TITLE

JOURNAL

COMMENT

Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Aizawa, K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagakawa, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermosensitization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1. 275
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 10 days neonate
head"
/sex="mixed"
/tissue_type="head"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
transcribed by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAAATTAATCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:

BASE COUNT 79 a 55 c 81 g 60 t
ORIGIN

alignment_scores:

Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BB575989 ..

Align seg 1/1 to: BB575989 from: 1 to: 275

169 glyluserglylserThrlenu 178
|||||

5 GGGGAATCCGAAAGGCAACACACTC 34

seq_name: gb_est1:AM401018

seq_documentation_block:

LOCUS AM401018 389 bp mRNA EST 07-FEB-2000
DEFINITION LamidGest148est L.digitata gametophyte Lambda ZapII Laminaria
digitata cDNA similar to similar to zinc metalloprotease, mRNA
sequence.

ACCESSION AM401018
VERSION AM401018.1 GI:6919508

KEYWORDS EST.
SOURCE Laminaria digitata.
ORGANISM Eukaryota; stramenopiles; Phaeophyceae; Laminariales; Laminariaceae

REFERENCE 1 (bases 1 to 389)
AUTHORS Crepinheu,F., Roscoe,T., Kaas,R., Kioareg,B. and Boyen,C.
TITLE Characterisation of complementary DNAs from the Expressed Sequence
Tag analysis of life cycle stages of Laminaria digitata
(Phaeophyceae)
Unpublished (2000)

JOURNAL Contact: Boyen C
Centre d'Etudes Oceanologique et de Biologie Marine
CNRS-LPR9042, Universite P. & M. Curie
BP74, F-29682 Roscoff cedex, France
Tel: 33 2 98 29 23 32
Fax: 33 2 98 29 23 24
Email: esteb-roscoff.fr.

FEATURES
Location/Qualifiers

SOURCE

1..389
/organism="Laminaria digitata"
/db_xref="taxon:80365"
/clone_lib="L.digitata gametophyte Lambda ZapII"
/dev_stage="gametophyte"
/lab_host="SOLR"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
XhoI; The Laminaria gametophyte library, constructed by T
Roscoe and F Crepinheu, was oligo-(dtr) primed and
directionally cloned into a Uni-ZAPm XR vector
(Stratagene, La Jolla, CA, USA) using total mRNA from
gametophytes provided by R Kaas, IREMER Nantes, France."

BASE COUNT 90 a 94 c 120 g 85 t
ORIGIN

alignment_scores:

Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AM401018 ..

Align seg 1/1 to: AM401018 from: 1 to: 389

485 SerThyTyrSerLeuLeuArgTyrThr 494
|||||
211 TCAACGATTTCTTCGCTCGCATACACG 240

seq_name: gb_est1:AI222422

seq_documentation_block:

LOCUS AI222422 446 bp mRNA EST 30-NOV-1998
DEFINITION gh04f06.x1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
IMAGE:1843715 3', mRNA sequence.

ACCESSION AI222422
VERSION AI222422.1 GI:3804625

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 446)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 546 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 437.

FEATURES
Location/Qualifiers

1..446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:1843715"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19W, testis NHT, and B-cell
NCL-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 128 a 95 c 86 g 137 t
ORIGIN

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AI222422 ..

Align seg 1/1 to: AI222422 from: 1 to: 446

89 LeuphehISGlnThrSerGIuGIyAspleu 98
|||||
417 CTTTTCATCAGACATCAGAAAGGACTTG 446

seq_name: gb_est2:BF185320

seq_documentation_block:

LOCUS BF185320 826 bp mRNA EST 31-OCT-2000
DEFINITION 601844132P1 NIH-MGC_54 Homo sapiens cDNA clone IMAGE:4064886 5',
mRNA sequence.

ACCESSION BF185320
VERSION BF185320.1 GI:11063763

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 826)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:
 http://image.lnl.gov
 Plate: LICM899 row: p column: 07
 High quality sequence stop: 508.
 Location/Qualifiers

FEATURES

source

1. 826
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4064886"
 /clone_lib="NIH_MGC_54"
 /tissue_type="from chronic myelogenous leukemia"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
 Site.1: SfiI (ggccgctcgcc); Site.2: SfiI (ggccatagccg)
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-ATCTAGAGCGCGCGCGCATG-AT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT

180 a 218 c 252 g 176 t

ORIGIN

alignment_scores:

Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BF185320 ..

Align seg 1/1 to: BF185320 from: 1 to: 826

616 PhetYrGlyGlyAlaMetAlaSerTrpGlu 625
 ||||||||||||||||||||||||||||
 583 TTCTACGGGGGGGCAATGCGATCGTGGGAG 612

seq_name: gb_est1:AW307272

seq_documentation_block:

LOCUS AW307272 54 bp mRNA EST 20-JAN-2000
 DEFINITION sf54H07.y1 Gm-cl009 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl009-3830 5', mRNA sequence.

ACCESSION AW307272
 VERSION AW307272.1 GI:6719625

KEYWORDS

EST.
 soybean.
 SOURCE

ORGANISM

Glycine max
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
 Rosidae: eustosids I: Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

1 (bases 1 to 54)

REFERENCE

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Reising, B., Allen, M., Bowers
 Y., Person, B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project

TITLE

JOURNAL

COMMENT

Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World

Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Glibco.

FEATURES

source

1. 54
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl009-3830"
 /clone_lib="Gm-cl009"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II Xr; Site.1: EcoRI; Site.2:
 XhoI. The mRNA was isolated from entire roots of
 2-month-old 'Williams' plants that were greenhouse grown
 in 5-gallon pots. To suppress nodulation, Black Gold
 All-Purpose potting soil was supplemented with: 0.36g/L
 available phosphoric acid (P205), 20mg/L urea N, 0.16g/L
 S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.53mg/L Mn,
 0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following
 nutrients in a slow-release form (Osmocote): 0.165g/L
 ammonia N, 0.185g/L nitrate N, 0.35g/L available
 phosphoric acid, and 0.35g/L soluble potash. No nodules
 were visible on the roots at harvest. Stragene's cDNA
 synthesis kit (catalog #200401) was used to synthesize the
 cDNA. First-strand synthesis was performed with 5-methyl
 dCTP, hence the ligated cDNA is hemimethylated.
 Stragene's first-strand synthesis primer was used
 (GAGAGAGAGAGAGAGAGACTGCTCGAG(T)-18). After
 second-strand synthesis is, the cDNA ends were 'polished'
 with clone Pfu DNA polymerase, ligated to EcoRI adaptors,
 and phosphorylated. The XhoI site within the first-strand
 synthesis primer was restricted by digestion with XhoI;
 all XhoI sites in the cDNA would be protected by their
 hemimethylated status. The cDNA constructs were
 size-fractionated with a 400bp cutoff, using a SizeSep 400
 spun column from Pharmacia. The cDNA eluent was then
 ligated into Stragene's pBluescript II Xr predigested
 vector (pBluescript II SK(+)) that had been digested with
 EcoRI and XhoI, and phosphorylated. Both the white and
 blue colonies appear to contain recombinant plasmids with
 cDNA inserts. This library was constructed by Dr. Paul
 Keim and Dr. Virginia Coryell."

BASE COUNT

11 a 6 c 11 g 26 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AW307272 ..

Align seg 1/1 to: AW307272 from: 1 to: 54

891 AspaValGInGlySerLeuSerSerLeu 899
 ||||||||||||||||||||||||
 21 GATGTGCAAGGTTCTTATCTTCCTTA 47

seq_name: gb_est1:AU166132

seq_documentation_block:

LOCUS AU166132 187 bp mRNA EST 08-DEC-2000
 DEFINITION AU166132 Rice panicle shorter than 3cm Oryza sativa cDNA clone
 E31777, mRNA sequence.

ACCESSION AU166132
 VERSION AU166132.1 GI:11610456

KEYWORDS

EST.
 Oryza sativa.
 SOURCE

ORGANISM

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Erihartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 187)
 AUTHORS Sasaki,T. and Yamamoto,K.
 TITLE Rice cDNA from panicle (2000)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tcsasaki@agr.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/
 PROJECT = "RGP".
 E31777_102.

FEATURES
 source 1..187
 Location/Qualifiers
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E31777"
 /clone_lib="Rice panicle shorter than 3cm"
 /dev_stage="shorter than 3cm"
 /note="Organ: panicle"
 BASE COUNT 53 a 31 c 35 g 68 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AU16132/rev ..

Align seg 1/1 to reverse of: AU16132 from: 1 to: 187

355 HASSerHsThGInThrThLeuphe 363
 |||||||||||||||||||||||||
 103 CACTCACACACACACACATATTTC 77

seq_name: gb_est1:AM326139

seq_documentation_block:
 LOCUS AM326139 217 bp mRNA EST 25-APR-2001
 DEFINITION 18288 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AM326139
 VERSION AM326139.1 GI:6762060
 KEYWORDS EST.
 SOURCE
 ORGANISM
 cow.
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 217)
 Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G.,
 Petrea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 20

and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTACAGACG
 Plate: 12 row: M column: 2
 Seq primer: ATTTCGTGACACTATGAC.

FEATURES
 source 1..217
 Location/Qualifiers
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="vector: PCMV SPORT6; site_1: XbaI; site_2: XhoI;
 library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."
 BASE COUNT 35 a 74 c 70 g 38 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AM326139/rev ..

Align seg 1/1 to reverse of: AM326139 from: 1 to: 217

848 Vallysleuserlleusapleuser 856
 |||||||||||||||||||||||||
 207 GTGAGCTGTCTGACACTCTCC 181

seq_name: gb_est1:AM458474

seq_documentation_block:
 LOCUS AM458474 245 bp mRNA EST 17-JUL-2000
 DEFINITION sh09f07.y1 Gm-c1016 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-c1016-4286 5', mRNA sequence.
 ACCESSION AM458474
 VERSION AM458474.1 GI:7028691
 KEYWORDS EST.
 SOURCE
 ORGANISM
 soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 245)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,
 A., Bolla,B., Maria,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,
 Y., Peterson,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,
 R., Waterston,R. and Willson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Insert length: 942 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 228.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 295)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Riklin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The Mashu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 5, 1995 this sequence version replaced gi:797756.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1871
High quality sequence stops: 104 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1871 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 104.
Location/Qualifiers
1. .295
/organism="Homo sapiens"
/db_xref="GDB:402653"
/db_xref="taxon:9606"
/clone="IMAGE:30306"
/clone_lib="Soares Infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site: 1: Not I; Site 2: Hind III; 1st strand cDNA was primed with a Not I - 01190(dt) primer [5];
AAGTCGAGAAATTCGGCGCGCGAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the latmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 69 a 60 c 69 g 95 t 2 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x R40140 ..
Align seg 1/1 to: R40140 from: 1 to: 295
755 LeuThrasperleuglyasleuLys 763
|||||
11 TTAACCTGACTCTCTGTGGAACCTGAGC 37
seq_name: gb_est1:AI969370
seq_documentation_block:
LOCUS AI969370 304 bp mRNA EST 08-MAR-2000
DEFINITION wg91f05.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2480001.3,
similar to SW:CEBD_HUMAN P49716 CCAAT/ENHANCER BINDING PROTEIN
DELTA ;, mRNA sequence.
ACCESSION AI969370
VERSION AI969370.1 GI:5766188
KEYWORDS EST,
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 304)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality
Insert Length: 828 Std Error: 0.00
Seq primer: -400p from Glbco
High quality sequence stop: 1.
Location/Qualifiers
1. 304
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2480001"
/clone_lib="NCI CGAP GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT773b-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1257096-1258631, 1469064-1470963, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 53 a 98 c 91 g 62 t
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AI969370 ..
Align seg 1/1 to: AI969370 from: 1 to: 304
146 HishisArgValGluGlnLeuThrIleu 154
|||||
197 CACCATCGCTGTGAGCAGCTGACGCTG 223
seq_name: gb_est1:BE424128
seq_documentation_block:
LOCUS BE424128 314 bp mRNA EST 24-JUL-2000
DEFINITION WHE0076_E10_120ZS Wheat endosperm cDNA library Triticum aestivum
CDNA clone WHE0076_E10_120, mRNA sequence.
ACCESSION BE424128
VERSION BE424128.1 GI:9421971
KEYWORDS EST,
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
; Triticaceae; Triticum.

REFERENCE 1 (bases 1 to 314)
 AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Galli,G., Han,P.S., Hsia
 ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and
 Tong,J.C.
 TITLE The structure and function of the expressed portion of the wheat
 genomes - Endosperm cDNA library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@w.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Striatagene SK primer.
 FEATURES
 source Location/Qualifiers
 1..314
 /organism="Triticum aestivum"
 /cultivar="Cheyenne"
 /db_xref="taxon:4565"
 /clone="WHE0076_E10_I20"
 /clone_lib="Wheat endosperm cDNA library"
 /tissue_type="Endosperm"
 /dev_stage="5 to 30 days post anthesis seed"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda ZAP II, excised phagemid. Site 1:
 EcoRI. Seeds collected, endosperm isolated, and RNA
 prepared by Susan Altenbach. Library constructed by
 Striatagene, Inc. Plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab."
 BASE COUNT 82 a 62 c 68 g 102 t
 ORIGIN
 alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x BE424128 ..
 Align seg 1/1 to: BE424128 from: 1 to: 314
 451 AAlaGlyATGATGLeuSerSerLeuLeu 459
 ||||||||||||||||||||
 60 GCGCGAAGAGCGCTTTCATCGCTGCTT 86
 seq_name: gb_est2:BI281383
 seq_documentation_block:
 LOCUS BI281383 320 bp mRNA EST 19-JUL-2001
 DEFINITION UT-R-DC0-bz1-d-08-0-UT.s1 UT-R-DC0 Rattus norvegicus cDNA clone
 UT-R-DC0-bz1-d-08-0-UT 3', mRNA sequence.
 ACCESSION BI281383
 VERSION BI281383.1 GI:14931055
 KEYWORDS EST.
 SOURCE Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 320)
 REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 AUTHORS Normalization and subtraction: two approaches to facilitate gene
 TITLE discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a Bonaldo poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized seminal vesicles library cDNA Library Preparation:
 M.B. Soares Lab Clone distribution: clones will be available
 through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.
 FEATURES
 source Location/Qualifiers
 1..320
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UT-R-DC0-bz1-d-08-0-UT"
 /clone_lib="UT-R-DC0"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTZ19-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; The UT-R-DC0
 library is a non-normalized library constructed from rat
 seminal vesicle tissue. For a detailed description of the
 library from which this clone was derived, please visit
 our web site at ratest.eng.uiowa.edu. The subtraction has
 been previously described in (Bonaldo, Lennon and Soares,
 Genome Research 6:791-806, 1996)
 TAG_LIB=UT-R-DC0
 TAG_TISSUE=seminal vesicles
 TAG_SEQ=GTGATTACAC"
 BASE COUNT 65 a 51 c 72 g 132 t
 ORIGIN
 alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x BI281383/rev ..
 Align seg 1/1 to reverse of: BI281383 from: 1 to: 320
 1006 GAlpheAsPASPAsPleuSerVal 1014
 ||||||||||||||||||||
 161 CAATTGACGATGATGACCTCTCTGTA 135
 seq_name: gb_est2:BI280717
 seq_documentation_block:
 LOCUS BI280717 346 bp mRNA EST 19-JUL-2001
 DEFINITION UT-R-DC0-bz1-a-06-0-UT.s1 UT-R-DC0 Rattus norvegicus cDNA clone
 UT-R-DC0-bz1-a-06-0-UT 3', mRNA sequence.
 ACCESSION BI280717
 VERSION BI280717.1 GI:14929761
 KEYWORDS EST.
 SOURCE Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 346)
 REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 AUTHORS Normalization and subtraction: two approaches to facilitate gene
 TITLE discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized seminal vesicles library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
1..346
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DC0-b21-a-06-0-UI"
/clone_lib="UI-R-DC0"
/dev_strage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0 library is a non-normalized library constructed from rat seminal vesicle tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-DC0
TAG_TISSUE=seminal vesicles
TAG_SEQ=GTGATTACAC"

BASE COUNT 72 a 52 c 81 g 140 t 1 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BI280717/rev ..
Align seg 1/1 to reverse of: BI280717 from: 1 to: 346

1006 GlnPhetapsPaspaspleuServal 1014
|||||
152 CAATTTCACGATGATGACCTCTCTGTA 126

seq_name: gb_est1:AI436226

seq_documentation_block:
LOCUS AI436226 349 bp mRNA EST 13-APR-1999
DEFINITION th78h09.x1 Soares.NHMPu_S1 Homo sapiens cDNA clone IMAGE:2124833
3', mRNA sequence.
ACCESSION AI436226
VERSION AI436226.1 GI:4308704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 349)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 846 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 273.

FEATURES
source
1..349
Location/Qualifiers
1..349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2124833"
/clone_lib="Soares.NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of Plasmid DNA from three normalized libraries (melanocyte 2NbM, pregnant uterus 2NbM, and fetal heart 2NbM) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 112 a 65 c 56 g 116 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AI436226 ..
Align seg 1/1 to: AI436226 from: 1 to: 349

700 ValAlaGlySerLeuValLeu 708
|||||
209 GTTGCTGAAGCCCTCTCTCGTTTG 235

seq_name: gb_est1:AI384131

seq_documentation_block:
LOCUS AI384131 365 bp mRNA EST 28-MAR-1999
DEFINITION te36h05.x1 Soares.NHMPu_S1 Homo sapiens cDNA clone IMAGE:2088825
3', mRNA sequence.
ACCESSION AI384131
VERSION AI384131.1 GI:4196912
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 365)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1014 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 365.

FEATURES
source
1..365
Location/Qualifiers
1..365
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

/clone_1lb="IMAGE:2088825"
/dev_stage="Soares_NhMPu_S1"
/lab_host="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/clone="DHI08"
/notes="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBH, pregnant uterus
NBHPU, and fetal heart NBH13H) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT      117 a      67 c      59 g      122 t
ORIGIN

alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AI384131 ..

Align seg 1/1 to: AI384131 from: 1 to: 365

700 ValaGlySerLeuSerLeuValLeu 708
|||||
205 GTTCTGCAAGCCTTCTCTGTTTG 231

seq_name: gb_est1:AA157463

seq_documentation_block:
LOCUS      AA157463      382 bp      mRNA      EST      11-DEC-1996
DEFINITION z650405.r1 StrataGene endothelial cell 937223 Homo sapiens CDNA
ACCESSION  AA157463
VERSION     AA157463.1 GI:1729070
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 382)
AUTHORS   Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevasakis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Merra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
JOURNAL    MEDLINE
COMMENT    Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28M13 rev2 from Amer sham
High quality sequence stop: 323.
Location/Qualifiers
1. 382
/organism="Homo sapiens"
/db_xref="GDB:4621699"
/db_xref="taxon:9606"
/clone="IMAGE:590313"

```

```

/clone_1lb="StrataGene endothelial cell 937223"
/dev_stage="umbilical vein, 1 passage"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site 1: EcoRI, Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Umbilical vein endothelial cells, passaged once. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5'
CTGACGTTTTTTTTTTTTTTT 3'"

BASE COUNT      124 a      68 c      73 g      116 t      1 others
ORIGIN

alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA157463/rev ..

Align seg 1/1 to reverse of: AA157463 from: 1 to: 382

700 ValaGlySerLeuSerLeuValLeu 708
|||||
277 GTTCTGCAAGCCTTCTCTGTTTG 251

seq_name: gb_est1:AI348317

seq_documentation_block:
LOCUS      AI348317      386 bp      mRNA      EST      01-FEB-1999
DEFINITION G015a06.x1 NCI-CGAP_Lu5 Homo sapiens CDNA clone IMAGE:1908562 3',
mRNA sequence.
ACCESSION  AI348317
VERSION     AI348317.1 GI:4085511
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 386)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/db/rdp/image/image.html
Insert length: 808 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 361.
Location/Qualifiers
1. 386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1908562"
/clone_1lb="NCI-CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DHI08"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - Oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was

```

BASE COUNT 124 a 69 c 64 g 129 t
 ORIGIN constructed by Bento Soares and M. Fatima Bonaldo. "

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AI348317 ..

Align seg 1/1 to: AI348317 from: 1 to: 386

700 VALAAGlySerLeuSerLeuValLeu 708
 ||||||||||||||||||||||||||||
 206 GTTGCTGGAAGCCTTCTCTGTTTGG 232

seq_name: gb_est2:BE862922

seq_documentation_block:
 LOCUS BE862922 390 bp mRNA EST 29-SEP-2000
 DEFINITION UI-M-BG1-aif-g-09-0-UI.F1 NIH_BMAP_MSC_N Mus musculus cDNA clone
 UI-M-BG1-aif-g-09-0-UI 5', mRNA sequence.
 ACCESSION BE862922
 VERSION BE862922.1 GI:10382372
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 390)
 REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (?), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 CDNA Library Preparation: M.B. Soares Lab Clone distribution:
 Researchers may obtain BMAP CDNA clones from RESEARCH GENETICS. It
 should be noted that Bento Soares is generating a small number of
 additional specialized non-redundant arrays of BMAP cDNAs whose
 availability will be considered under appropriate and limited
 collaborative arrangements
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..390
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BG1-aif-g-09-0-UI"
 /clone_1ib="NIH_BMAP_MSC_N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p7773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_MSC_N library is a normalized library constructed
 from mouse spinal cord. The tag is a string of 5
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 '1996. Tissue provided by Ms. Anne Novakovich,
 Zivic-Miller Laboratories."

BASE COUNT 137 a 81 c 110 g 62 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x BE862922 ..

Align seg 1/1 to: BE862922 from: 1 to: 390

746 LeuGlnAsnGlnArgLeuProGlyGly 754
 ||||||||||||||||||||||||||||
 56 CTGCAAAATCAAGCGTACTGTGAGGA 82

seq_name: gb_est1:AA865992

seq_documentation_block:
 LOCUS AA865992 392 bp mRNA EST 12-MAR-1998
 DEFINITION oh29g09.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1459264 3'
 similar to gb:U02389 MITOCHONDRIAL LON PROTEASE HOMOLOG PRECURSOR
 (HUMAN);, mRNA sequence.
 ACCESSION AA865992
 VERSION AA865992.1 GI:2958268
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 392)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@email.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbtrp/image/image.html

Trace considered overall poor quality
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1..392
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1459264"
 /clone_1ib="NCI_CGAP_Kid6"
 /sex="mixed"
 /tissue_type="Kidney tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: Kidney; Vector: Bluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
 GAATTCGCGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTGTGTTTTTTTTTTT 3' Average Insert size: 1.0 kb."

BASE COUNT 73 a 117 c 113 g 89 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AA865992/rev ..

Align seg 1/1 to reverse of: AA865992 from: 1 to: 392

450 ThrAlaGlyArgArgLeuSerLeu 458
 |||||||
 180 ACTGACGCGCCGATGTCAAGCCTG 154

seq_name: gb_est2:BF940899

seq_documentation_block:

LOCUS BF940899 399 bp mRNA EST 22-JAN-2001
 DEFINITION nu61a10.x1 NCI_CGAP_Brn41 Homo sapiens cDNA clone IMAGE:3174427 3',
 mRNA sequence.

ACCESSION BF940899
 VERSION BF940899.1 GI:12358219

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 399)

AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BTCAP), Tumor Gene Index

JOURNAL Unpublished (1998)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Robert Jenkins, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D. and M.Fatima
 Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
 info@image.llnl.gov
 Seq primer: -40UP from GIBCO.

FEATURES
 source Location/Qualifiers

1..399
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3174427"
 /clone_id="NCI_CGAP_Brn41"
 /tissue_type="oligodendroglioma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."

BASE COUNT 128 a 72 c 67 g 131 t 1 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BF940899 ..

Align seg 1/1 to: BF940899 from: 1 to: 399

700 ValAlaGlySerLeuSerLeuValLeu 708
 |||||||
 202 GTTGCTGGAGACCTTCTCTGGGTTTG 228

seq_name: gb_est2:R42493

seq_documentation_block:

LOCUS R42493 400 bp mRNA EST 22-MAY-1995
 DEFINITION y902908.s1 Soares infant brain INIB Homo sapiens cDNA clone
 IMAGE:31060 3', mRNA sequence.

ACCESSION R42493
 VERSION R42493.1 GI:817257

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 400)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
 M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston,
 R., Williamson, A., Wohlmann, P. and Wilson, R.

JOURNAL The WashU-Merck EST Project
 Unpublished (1995)

COMMENT On May 8, 1995 this sequence version replaced gi:800717.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert size: 1919

High quality sequence stops: 311
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1919 Std Error: 0.00
 Seq primer: Promega -21m13
 High quality sequence stop: 311.

FEATURES
 source Location/Qualifiers

1..400
 /organism="Homo sapiens"
 /db_xref="GDB:403407"
 /db_xref="taxon:9606"
 /clone_image="31060"
 /clone_id="Soares infant brain INIB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
 I; Site_2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5'
 AACTGGAAGATTCGCGCGCGAGATTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the lafmid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 97 a 82 c 96 g 123 t 2 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x R42493 ..

Align seg 1/1 to: R42493 from: 1 to: 400

755 LeuThrAspSerLeuGlyAsnLeuLys 763
 |||||||
 11 TTAAGTACTCTCTTGGAACCTTGAAG 37

seq_name: gb_est1:A1191648

```

seq_documentation_block:
LOCUS      A1191648      404 bp      mRNA      EST      28-OCT-1998
DEFINITION gq47db07.x1 Soares fetal heart NBH19M Homo sapiens cDNA clone
IMAGE:1732597 3' similar to contains Alu repetitive element; mRNA
sequence.
ACCESSION  A1191648
VERSION     A1191648.1  GI:3742857
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 404)
AUTHORS     NC-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rt@mail.nih.gov
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert length: 554 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 370.
FEATURES
    source
        1..404
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:1732597"
        /clone_lib="Soares_fetal_heart_NBH19M"
        /sex="unknown"
        /dev_stage="19 weeks"
        /lab_host="DH10B (ampicillin resistant)"
        /note="Organ: heart; Vector: pT73D (Pharmacia) with a
        modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
        strand cDNA was primed with a Not I - oligo(dT) primer [5'
        TGTTACCATCTGAGTGGAGCGCCCATCTTTTCTTTTCTTTT 3'],
        double-stranded cDNA was size selected, ligated to Eco RI
        adapters (Pharmacia), digested with Not I and cloned into
        the Not I and Eco RI sites of a modified pT73 vector
        (Pharmacia). Library went through one round of
        normalization to a Cot = 5. Library constructed by
        M.Falima Bonaldo. This library was constructed from the
        same fetus as the fetal lung library, Soares fetal lung
        NBH19M."
BASE COUNT      85 a      130 c      93 g      96 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x A1191648 ..
Align seg 1/1 to: A1191648 from: 1 to: 404
450 ThrAlaGlyArgArgLeuSerSerLeu 458
|||||
303 ACTGCTGTCGAGAGGCTCTCTCCCTA 329
seq_name: gb_gss:AQ889169
seq_documentation_block:
LOCUS      AQ889169      404 bp      DNA      GSS      10-NOV-1999
DEFINITION HS_2161.B1.A01_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.
ACCESSION  AQ889169
VERSION     AQ889169.1  GI:6345359
KEYWORDS    GSS..
SOURCE      human.

```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 404)
AUTHORS       Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
            Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
            Hood L.
TITLE         Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT       Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.husc.washington.edu
            Plate: 2161 Row: B Column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 404.
FEATURES
    source
        1..404
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="Plate=2161 Col=1 Row=B"
        /clone_lib="CIT Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
        E-Coli DH10B"
BASE COUNT      119 a      93 c      70 g      122 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AQ889169/rev ..
Align seg 1/1 to reverse of: AQ889169 from: 1 to: 404
52 ArgGlyIleIleHsMetIleLeuLys 60
|||||
199 AGAGGGATCATTCACATGATTTTGAAA 173
seq_name: gb_est:AL509287
seq_documentation_block:
LOCUS      AL509287      432 bp      mRNA      EST      04-JAN-2001
DEFINITION AL509287 Hordeum vulgare Barke developing carypsps (3.-15.DAP)
            Hordeum vulgare cDNA clone HY01E24u 3', mRNA sequence.
ACCESSION  AL509287
VERSION     AL509287.1  GI:12035790
KEYWORDS    EST.
SOURCE      barley.
ORGANISM    Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE     1 (bases 1 to 432)
AUTHORS       Michalek W., Weschke W., Pleissner K.-P. and Graner A.
TITLE         EST sequencing and analysis in barley
JOURNAL       Unpublished (2000)
COMMENT       Contact: Michalek W
            Institute for Plant Genetics and Crop Plant Research
            Corrensstr. 3, D-06466 Gatersleben, Germany
            Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de

```

FEATURES Seq primer: M13uni primer for 3' end.
Location/Qualifiers
SOURCE 1. 432
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY01E24u"
/clone.lib="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"
/tissue_type="developing caryopsis (3.-15.DAP)"
/lab_host="X10LR"
/note="Vector: plasmid pBK-CMV. Site_1: EcoRI; Site_2: XhoI. mRNA was made from developing caryopsis (3.-15.DAP) of spring barley variety 'Barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb. Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'- and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp."

BASE COUNT 111 a 96 c 100 g 125 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AL509287/rev ..

Align seg 1/1 to reverse of: AL509287 from: 1 to: 432

451 AlAGlYARgARgLeuSerLeu 459
|||||
214 GCGGAGAGAGGCTTCATCGCTCTT 188

seq_name: gb_est1:AW266497

seq_documentation_block: 432 bp mRNA EST 30-DEC-1999
LOCUS AW266497
DEFINITION YAS1 gastric carcinoma cell GC7901 Homo sapiens cDNA, mRNA
sequence.
ACCESSION AW266497
VERSION AW266497.1 GI:6646675
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS Zhao,J.R., Yan,X.J., Han,F.C., Cui,D.X., Hou,Y., Yan,Q.J. and Su
C.Z.
TITLE Gastric associated differentially expressed gene mRNA sequence
JOURNAL Unpublished (1999)
COMMENT Contact: Zhao JR
Institute of Genetic Diagnosis, the Fourth Military Medical
University
Xi'an, China, 710032
Tel: 86 029 3374771
Email: zhaojr@263.net
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
SOURCE 1. 432
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="gastric carcinoma cell GC7901"
/cell_type="gastric carcinoma cell GC7901"
/note="Differentially expressed gene sequences from human

gastric carcinoma cell GC7901. The method used is
DDRT-PCR. The control cell line is gastric epithelial cell
GES-1"

BASE COUNT 200 a 42 c 59 g 131 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AW266497 ..

Align seg 1/1 to: AW266497 from: 1 to: 432

169 GLYGLUserGlySGLySSerThr 177
|||||
5 GCGCATCAGCTAAGCAATCATCA 31

seq_name: gb_est2:BG159456

seq_documentation_block: 436 bp mRNA EST 06-FEB-2001
LOCUS BG159456
DEFINITION OV2.8.D03.g1.A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BG159456
VERSION BG159456.1 GI:12693120
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 436)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: PolYTmIX
High quality sequence start: 46
High quality sequence stop: 433
POLYA=No.

FEATURES
SOURCE 1. 436
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone.lib="Ovary 2 (OV2)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass exsiston."

BASE COUNT 133 a 109 c 121 g 73 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BG159456 ..

Align seg 1/1 to: BG159456 from: 1 to: 436

259 ProGUinleGlualaleuIlyeSglu 267
 131 CCGAGATAGACGACTGATCAAGAG 157

seq_name: gb_est1:AL510027

seq_documentation_block:

LOCUS AL510027 439 bp mRNA EST 04-JAN-2001
 DEFINITION Hordeum vulgare Barke developing caryopsis (3'-15.DAP)
 Hordeum vulgare cDNA clone HY03K06u 3', mRNA sequence.

ACCESSION AL510027
 VERSION AL510027.1 GI:12036530
 KEYWORDS EST

SOURCE
 ORGANISM

Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 Triticeae; Hordeum.

REFERENCE
 AUTHORS Michael M., Weschke, W., Pleissner, K.-P. and Graner, A.
 TITLE EST sequencing and analysis in barley
 JOURNAL Unpublished (2000)

COMMENT
 Contact: Michael M

Institute for Plant Genetics and Crop Plant Research
 Corrensstr.3, D-06466 Gatersleben, Germany
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
 Seq primer: M13uni primer for 3' end.

FEATURES
 SOURCE

Location/Qualifiers

1..439
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HY03K06u"
 /clone_lib="Hordeum vulgare Barke developing caryopsis
 (3'-15.DAP)"
 /tissue_type="developing caryopsis (3'-15.DAP)"
 /lab_host="X10LR"

/note="Vector: plasmid pBK-CMV, Site_1: EcoRI; Site_2:
 XhoI; mRNA was made from developing caryopsis (3'-15.DAP)
 of spring barley variety 'Barke', a high quality malting
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
 (3'-end of cDNA). NOTE: Due to a cloning artefact caused
 by the kit, in most cases the EcoRI site is NOT present,
 as well as the EcoRI adapter. Average insert size is 1 kb
 trimmed from the 5'-and 3'-end until a 50 bp window
 contains less than two ambiguities. The maximum length was
 set to 700 bp"

BASE COUNT 111 a 97 c 101 g 130 t
 ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AL510027/rev ..

Align seg 1/1 to reverse of: AL510027 from: 1 to: 439

451 AAlaGlyArgLeuSerLeuValLeu 459

218 GCGGAGAGAGAGCCTTTCCTGCTT 192

seq_name: gb_est1:AI813286

seq_documentation_block:

LOCUS AI813286 444 bp mRNA EST 21-DEC-1999
 DEFINITION wJ06c12.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2402038 3',

ACCESSION mRNA sequence.

VERSION AI813286.1 GI:5424501

KEYWORDS EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Robert Strausberg, Ph.D.

Email: cga@b-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.

Emmett-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.lnl.gov/bbrp/image/image.html

Insert length: 644 Std Error: 0.00

Seq primer: -400p from Gibco

High quality sequence stop: 424.

Location/Qualifiers

1..444

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2402038"

/clone_lib="NCI_CGAP_Kid12"

/tissue_type="2 pooled tumors (clear cell type)"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Plasmid DNA from the normalized library NCI_CGAP_Kid12 was

prepared, and ss circles were made in vitro. Following RAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(clones 1323912-1325831, 1471368-1472903 and

1492104-1493255). Subtraction by Bento Soares and M.

BASE COUNT 139 a 73 c 72 g 159 t 1 others
 ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AI813286 ..

Align seg 1/1 to: AI813286 from: 1 to: 444

700 ValAlaGlySerLeuSerLeuValLeu 708

366 GTTGCGGAGAGCCTTCTCTGCTTTTG 392

seq_name: gb_est1:AW149096

seq_documentation_block:

LOCUS AW149096 448 bp mRNA EST 03-NOV-1999
 DEFINITION xfl10b02.x1 NCI_CGAP_Kid8 Homo sapiens cDNA clone IMAGE:2617611 3',

ACCESSION AW149096

VERSION AW149096.1 GI:6196992

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 448)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdrp/image/image.html

FEATURES
 source
 1..448
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2617611"
 /clone_lib="NCI-CGAP_Kid8"
 /tissue_type="renal cell tumor"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.2 kb. Life Technologies catalog #: 11524-014"

BASE COUNT
 152 a 80 c 69 g 147 t

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AM149096 ..

Align seg 1/1 to: AM149096 from: 1 to: 448

700 VALAAGIySerLeuSerLeuValLeu 708
 ||||||||||||||||||||||||||||
 369 GTTGCTGAGAGCCTTCTGCTTTTG 395

seq_name: gb_gss:A2598362

seq_documentation_block:
 LOCUS A2598362 456 bp DNA GSS 13-DEC-2000
 DEFINITION M0413G07F Mouse 10kb plasmid UUGCIM library musculus genomic clone UUGCIM0413G07 F, DNA sequence.
 ACCESSION A2598362
 VERSION A2598362.1 GI:11720552
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 456)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss

University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0413 row: G column: 07
 Seg primer: CATTGTAACGACGCGCACT
 Class: plasmid ends
 High quality sequence stop: 456.
 location/Qualifiers
 1..456
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGCIM0413G07"
 /clone_lib="Mouse 10kb plasmid UUGCIM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: FMD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|g1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
 167 a 82 c 115 g 92 t

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x A2598362/rev ..

Align seg 1/1 to reverse of: A2598362 from: 1 to: 456

895 SerLeuSerLeuLeuHisLeu 903
 ||||||||||||||||||||||||||||
 216 TCCTTGCTCCTCACTTTAAAGCATCTT 190

seq_name: gb_est1:A1743742

seq_documentation_block:
 LOCUS A1743742 457 bp mRNA EST 19-DEC-1999
 DEFINITION w953803.x1 Soares_NSF_F8_9W_OT_PA_P-S1 Homo sapiens cDNA clone IMAGE:2368805 3', mRNA sequence.
 ACCESSION A1743742
 VERSION A1743742.1 GI:5112030
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 457)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (Info@image.llnl.gov) for further information.
 Insert Length: 1204 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 337.
 Location/Qualifiers

FEATURES
 source
 1..457
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2368805"
 /clone_lib="Soares NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and cloneids: Soares NBHSF pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HP-9W pool 1:
 758280-760583, 772104-774407 Soares NBHPA pool 1:
 304776-306311, 320136-322823, 326280-326653 Soares NBHOT
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 Soares 75 c 74 g 156 t 1 others

BASE COUNT 151 a 75 c 74 g 156 t 1 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AT743742 ..
 Align seg 1/1 to: AT743742 from: 1 to: 457

700 ValAlAGlySerLeuSerLeuValLeu 708
 |||||
 277 GTTGCTGGAAGCCTTCCTGCTGTTTG 303

seq_name: gb_est1:AL512027

seq_documentation_block:
 LOCUS AL512027 458 bp mRNA EST 04-JAN-2001
 DEFINITION AL512027 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
 Hordeum vulgare cDNA clone HY10C09u 3', mRNA sequence.
 ACCESSION AL512027
 VERSION AL512027.1 GI:12038530
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 1 (bases 1 to 458)
 Michael, W., Weschke, W., Pleisner, K.-P. and Graner, A.
 EST sequencing and analysis in barley
 Unpublished (2000)
 Contact: Michael W
 Institute for Plant Genetics and Crop Plant Research
 Corrensstr. 3, D-06466 Gatersleben, Germany
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de

Seq primer: M3uni primer for 3'end.
 Location/Qualifiers

FEATURES
 source
 1..458
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HY10C09u"
 /clone_lib="Hordeum vulgare Barke developing caryopsis
 (3.-15.DAP)"
 /tissue_type="developing caryopsis (3.-15.DAP)"
 /lab_host="XLOLR"
 /note="Vector: plasmid pBK-CMV, Site_1: EcoRI; Site_2:
 XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
 of spring barley variety 'Barke', a high quality malting
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
 (3'-end of cDNA). NOTE: Due to a cloning artefact caused
 by the kit, in most cases the EcoRI site is NOT present,
 as well as the EcoRI adapter. Average insert size is 1 kb
 Sequence trimming: Vector sequences and sequence ends were
 trimmed from the 5'- and 3'-end until a 50 bp window
 contains less than two ambiguities. The maximum length was
 set to 700 bp"

BASE COUNT 115 a 104 c 107 g 130 t 2 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AL512027/rev ..
 Align seg 1/1 to reverse of: AL512027 from: 1 to: 458

451 AlAGlyArgArgLeuSerLeuLeu 459
 |||||
 211 GCGGGAAGAGCCTTCATCGCTGCTT 185

seq_name: gb_est2:BF435974

seq_documentation_block:
 LOCUS BF435974 459 bp mRNA EST 30-MAR-2001
 DEFINITION nab75h07.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:3273829 3', mRNA sequence.
 ACCESSION BF435974
 VERSION BF435974.1 GI:11448289
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 459)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (Info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 443.
 Location/Qualifiers

FEATURES
 source
 1..459
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3273829"
 /clone_lib="Soares NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOF pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaudo."

BASE COUNT 129 a 80 c 75 g 175 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BF435974 ..

Align seg 1/1 to: BF435974 from: 1 to: 459

700 ValAlaGlySerLeuValIleu 708
|||||
242 GTGCTGGAGCCCTTCTCTGTTTG 268

seq_name: gb_est1:AM301228

seq_documentation_block:
LOCUS AM301228 461 bp mRNA EST 18-JAN-2000
DEFINITION xs69b02.x1 NCI-CGAP_Kid1 Homo sapiens cDNA clone IMAGE:2774859 3'
similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AM301228
VERSION AM301228.1 GI:6710905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 461)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrp/image/image.html
Seq primer: -40UP from Gibco

High quality sequence stop: 444.

FEATURES

Location/Qualifiers
1..461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2774859"
/clone_lib="NCI-CGAP_Kid1"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaudo."

BASE COUNT 154 a 76 c 74 g 157 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AM301228 ..

Align seg 1/1 to: AM301228 from: 1 to: 461

700 ValAlaGlySerLeuValIleu 708
|||||
276 GTGCTGGAGCCCTTCTCTGTTTG 302

seq_name: gb_est1:AA032702

seq_documentation_block:
LOCUS AA032702 466 bp mRNA EST 22-AUG-1996
DEFINITION m134h10.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:465475 5', mRNA sequence.

ACCESSION AA032702
VERSION AA032702.1 GI:1505294
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 466)
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

AUTHORS Mairra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, F., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

COMMENT The WashU-HMNI Mouse EST Project
Unpublished (1996)

JOURNAL Contact: Mairra W/Mouse EST Project
WashU-HMNI Mouse EST Project

WASHINGTON University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279291

Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 444.

Location/Qualifiers
1..466

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:465475"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"

/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TCTTACCAATCTGAGTCGAGCCGCCGCGAATAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 j; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 149 a 104 c 124 g 89 t

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AA032702 ..

Align seg 1/1 to: AA032702 from: 1 to: 466

746 LeuGlnAsnGlnArgLeuProGlyGly 754
 ||||||||||||||||||||||||||||
 383 CTGCAAAATCAAGGCTACTCGAGGCA 409

seq_name: gb_est2:BG055260

seq_documentation_block:
 LOCUS BG055260 469 bp mRNA EST 25-JAN-2001
 DEFINITION na52c01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3407208 3',
 mRNA sequence.
 ACCESSION BG055260
 VERSION BG055260.1 GI:12513540
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 469)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: 40up from Gibco
 High quality sequence stop: 462.
 Location/Qualifiers
 1..469
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3407208"
 /clone_lib="NCI_CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Plasmid DNA from the normalized
 library NCI_CGAP_Lu5 was prepared, and ss circles were
 made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (clonoids
 1414920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 81 c 78 g 161 t

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x BG055260 ..

Align seg 1/1 to: BG055260 from: 1 to: 469

700 ValAlaGlySerLeuSerLeuValLeu 708
 ||||||||||||||||||||||||||||
 210 GTTCTGTGAGAGCCCTTCTCTGTTTG 236

seq_name: gb_est1:BE517849

seq_documentation_block:
 LOCUS BE517849 480 bp mRNA EST 08-AUG-2000
 DEFINITION WHE0803_B04.D0725 wheat vernalized crown cDNA library Trilicium
 aestivum cDNA clone WHE0803_B04.D07, mRNA sequence.
 ACCESSION BE517849
 VERSION BE517849.1 GI:9741879
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Trilicium aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
 ; Triticeae; Trilicium.
 1 (bases 1 to 480)
 REFERENCE Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
 ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
 Seaton,C.I. and Tong,J.C.
 The structure and function of the expressed portion of the wheat
 genomes - Vernalized crown cDNA library
 Unpublished (2000)
 JOURNAL Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 51053595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.
 Location/Qualifiers
 1..480
 /organism="Trilicium aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE0803_B04.D07"
 /clone_lib="wheat vernalized crown cDNA library"
 /tissue_type="Crown tissue of seedling"
 /dev_stage="Five-week old seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI, Site 2: XhoI. Seeds were germinated and
 grown at 4 C for 5 weeks. The tissue, total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pluscript
 phagemids in the T7 Close lab (Choi, Close, Fenton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."

BASE COUNT 131 a 127 c 123 g 99 t

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BE517849 ..

Align seg 1/1 to: BE517849 from: 1 to: 480

451 AlaglyArgArgLeuSerSerLeu 459

|||||

seq_name: gb_est2:BI279982

seq_documentation_block:

LOCUS BI279982 484 bp mRNA EST 19-JUL-2001
DEFINITION UI-R-DC0-cda-g-03-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone

ACCESSION BI279982

VERSION BI279982.1 GI:14928328

KEYWORDS EST

SOURCE Norway rat.
Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 484)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

AUTHORS TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL GENOME RES. 6 (9), 791-806 (1996)

COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscores@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized seminal vesicles library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=yes.

FEATURES

source

1..484

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-DC0-cda-g-03-0-UI"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pRT3D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0

library is a non-normalized library constructed from rat

seminal vesicle tissue. For a detailed description of the

library from which this clone was derived, please visit

our web site at rategen.uiowa.edu. The subtraction has

been previously described in (Bonaldo, Lennon and Soares,

Genome Research 6:791-806, 1996)

TAG_LIB=UI-R-DC0

TAG_TISSUE=seminal vesicles

TAG_SEQ=GTGATTACAC

BASE COUNT 110 a 82 c 111 g 180 t 1 others

ORIGIN

alignment_scores:

Quality: 9.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BI279982/rev ..

Align seg 1/1 to reverse of: BI279982 from: 1 to: 484

1006 GinpheaspaspaspaspLeuSerVal 1014

|||||

seq_name: gb_est1:AL509297

seq_documentation_block:

LOCUS AL509297 488 bp mRNA EST 04-JAN-2001

DEFINITION Hordelum vulgare Barke developing caryopsis (3.-15.DAP)

ACCESSION AL509297

VERSION AL509297.1 GI:12035800

KEYWORDS EST

SOURCE barley.

ORGANISM Hordelum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

; Triticeae; Hordelum.

REFERENCE 1 (bases 1 to 488)

AUTHORS Michael,M., Wessche,W., Pleissner,K.-P. and Graner,A.

JOURNAL EST sequencing and analysis in barley

UNPUBLISHED (2000)

COMMENT Contact: Michael M

Institute for Plant Genetics and Crop Plant Research

Corrensstr.3, D-06466 Gatersleben, Germany

Email: michael@ipk-gatersleben.de <http://ipk-gatersleben.de>

Seq primer: M13umi primer for 3' end.

FEATURES

source

1..488

/organism="Hordelum vulgare"

/cultivar="Barke"

/db_xref="taxon:4513"

/clone="HY01P12u"

/clone_lib="Hordelum vulgare Barke developing caryopsis

(3.-15.DAP)"

/tissue_type="developing caryopsis (3.-15.DAP)"

/lab_host="XLOLR"

/note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:

XhoI; mRNA was made from developing caryopsis (3.-15.DAP)

of spring barley variety 'Barke', a high quality malting

variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI

(3'-end of cDNA). NOTE: Due to a cloning artefact caused

by the kit, in most cases the EcoRI site is NOT present,

as well as the EcoRI adapter. Average insert size is 1 kb

Sequence trimming: Vector sequences and sequence ends were

trimmed from the 5'- and 3'-end until a 50 bp window

contains less than two ambiguities. The maximum length was

set to 700 bp"

BASE COUNT 124 a 110 c 116 g 138 t

ORIGIN

alignment_scores:

Quality: 9.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AL509297/rev ..

Align seg 1/1 to reverse of: AL509297 from: 1 to: 488

451 AlaglyArgArgLeuSerSerLeu 459

```

|||||
211 GCGGGAAGAGCGCTTCATCGCTGCTT 185
seq_name: gb_est1:AA156049

seq_documentation_block:
LOCUS      AA156049      489 bp      mRNA      EST      11-DEC-1996
DEFINITION  zos0d05.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA
ACCESSION  AA156049
VERSION    AA156049
KEYWORDS   AA156049.1 GI:1727674
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 489)
AUTHORS   Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Treaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
JOURNAL
MEDLINE
COMMENT    Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estelw@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 327.
FEATURES
    source
        1..489
            /organism="Homo sapiens"
            /db_xref="GDB:4621699"
            /db_xref="taxon:9606"
            /clone="IMAGE:590313"
            /clone_lib="Stratagene endothelial cell 937223"
            /dev_stage="umbilical vein, 1 passage"
            /lab_host="SOLR (kanamycin resistant)"
            /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
            XhoI; Cloned unidirectionally. Primer: Oligo dT.
            Umbilical vein endothelial cells, passaged once. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
            sequence: 5' GAATTCGACGACGAC 3' -3' adaptor sequence: 5'
            CTCGAGTTTCTTTTCTTTTCTT 3'"
BASE COUNT      157 a      87 c      81 g      164 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AA156049 ..
Align seg 1/1 to: AA156049 from: 1 to: 489
700 ValAAGlySerLeuSerLeuValLeu 708
|||||
208 GTTCTGGAACCTTTCTCTGCTTTTG 234
seq_name: gb_est2:BI281408
seq_documentation_block:
LOCUS      BI281408      490 bp      mRNA      EST      19-JUL-2001

```

```

DEFINITION  UI-R-DC0-bzj-f-10-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
UI-R-DC0-bzj-f-10-0-UI 3', mRNA sequence.
ACCESSION  BI281408
VERSION    BI281408.1 GI:14931101
KEYWORDS   EST
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE  1 (bases 1 to 490)
AUTHORS   Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   discovery
MEDLINE   Genome Res. 6 (9), 791-806 (1996)
97044477
COMMENT    Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized seminal vesicles library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
FEATURES
    source
        1..490
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-DC0-bzj-f-10-0-UI"
            /clone_lib="UI-R-DC0"
            /dev_stage="ADULT"
            /lab_host="DH10B (Life Technologies)"
            /note="Vector: pYT73D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0
            library is a non-normalized library constructed from rat
            seminal vesicle tissue. For a detailed description of the
            library from which this clone was derived, please visit
            our web site at ratseq.eng.uiowa.edu. The subtraction has
            been previously described in (Bonaldo, Lennon and Soares,
            Genome Research 6:791-806, 1996)
            TAG_LIB=UI-R-DC0
            TAG_RISUSE=seminal vesicles
            TAG_SEO=GTGATTCACAC"
BASE COUNT      113 a      83 c      112 g      182 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x BI281408/rev ..
Align seg 1/1 to reverse of: BI281408 from: 1 to: 490
1006 GlnPheAspAspAspLeuSerVal 1014
|||||
472 CAATTGACGATGATGACCTCTGTGA 446
seq_name: gb_est1:AI859851

```

```

seq_documentation_block:
LOCUS      A1859851      491 bp      mRNA      EST      07-MAR-2000
DEFINITION  MNCI-CGAP ut4 Homo sapiens cDNA clone IMAGE:2436588 3'
              similar to contains Alu repetitive element, contains element MSRI
              repetitive element ;, mRNA sequence.
ACCESSION  A1859851
VERSION    A1859851.1 GI:5513456
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 491)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the J.M.A.G.E. Consortium/LNL at:
            www-bio.lnl.gov/bbrp/image/image.html
            Insert Length: 1146 Std Error: 0.00
            Seq primer: -400p from Glbco
            High quality sequence stop: 412.
            Location/Qualifiers
                source
                    1..491
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2436588"
                        /clone_id="MNCI-CGAP-ut4"
                        /tissue_type="serous papillary carcinoma, high grade, 2
                        pooled tumors"
                        /lab_host="DH10B"
                        /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                        Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                        Average insert size 1.48 kb. Life Technologies catalog #:
                        11542-016"
BASE COUNT      104 a      156 c      122 g      109 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x A1859851 ..
Align seg 1/1 to: A1859851 from: 1 to: 491
    450 ThTAlaGlyAArgArgLeuSerSerLeu 458
    11111111111111111111111111111111
    303 ACTGCTGCAGAGAGCTCTCTCCCTA 329
seq_name: gb_est2:B1280561
seq_documentation_block:
LOCUS      B1280561      491 bp      mRNA      EST      19-JUL-2001
DEFINITION  UI-R-DC0-bz1-h-11-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
              UI-R-DC0-bz1-h-11-0-UI 3', mRNA sequence.
ACCESSION  B1280561
VERSION    B1280561.1 GI:14929468
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE  1 (bases 1 to 491)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704447
COMMENT   Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dt track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dt track served to verify it as a clone from the
            non-normalized seminal vesicles library cDNA library preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=yes.
            Location/Qualifiers
                source
                    1..491
                        /organism="Rattus norvegicus"
                        /strain="Sprague-Dawley"
                        /db_xref="taxon:10116"
                        /clone="UI-R-DC0-bz1-h-11-0-UI"
                        /clone_id="UI-R-DC0"
                        /dev_stage="ADULT"
                        /lab_host="DH10B (Life Technologies)"
                        /note="Vector: pRT3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0
                        library is a non-normalized library constructed from rat
                        seminal vesicle tissue. For a detailed description of the
                        library from which this clone was derived, please visit
                        our web site at rates.eng.uiowa.edu. The subtraction has
                        been previously described in (Bonaldo, Lennon and Soares,
                        Genome Research 6:791-806, 1996)
                        TAG_Lib=UI-R-DC0
                        TAG_TISSUE=seminal vesicles
                        TAG_SEQ=GTGATTACAC"
BASE COUNT      112 a      83 c      113 g      183 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x B1280561/rev ..
Align seg 1/1 to reverse of: B1280561 from: 1 to: 491
    1006 GlnPhaSpaSpaSpaSpaSerSerVal 1014
    11111111111111111111111111111111
    475 CATTGTGACGATGATGACCTCTCTGTA 449
seq_name: gb_est2:B1280626
seq_documentation_block:
LOCUS      B1280626      491 bp      mRNA      EST      19-JUL-2001
DEFINITION  UI-R-DC0-bzm-f-07-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
              UI-R-DC0-bzm-f-07-0-UI 3', mRNA sequence.
ACCESSION  B1280626
VERSION    B1280626.1 GI:14929596
KEYWORDS   EST.
SOURCE     Norway rat.

```

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 491)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized seminal vesicles library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=Yes.

FEATURES
source location/Qualifiers
1..491
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DC0-bzm-f-07-0-UI"
/clone_1lb="UI-R-DC0"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTZ19D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0
library is a non-normalized library constructed from rat
seminal vesicle tissue. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_1lb="UI-R-DC0"
TAG_TISSUE="seminal vesicles"
TAG_SEQ="GTGATTACAC"
BASE COUNT 112 a 83 c 113 g 183 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BI280626/rev ..
Align seg 1/1 to reverse of: BI280626 from: 1 to: 491

1006 GlnPheAspAspAspPleuSerVal 1014
|||||
475 CAATTGACGATGATGACCTCTCTGTA 449

seq_name: gb_est2:BI280645

seq_documentation_block:
LOCUS BI280645 491 bp mRNA EST 19-JUL-2001
DEFINITION UI-R-DC0-bzm-h-02-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
ACCESSION BI280645
VERSION BI280645.1 GI:14929633

KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 491)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized seminal vesicles library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=Yes.

FEATURES
source location/Qualifiers
1..491
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DC0-bzm-h-02-0-UI"
/clone_1lb="UI-R-DC0"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTZ19D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0
library is a non-normalized library constructed from rat
seminal vesicle tissue. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_1lb="UI-R-DC0"
TAG_TISSUE="seminal vesicles"
TAG_SEQ="GTGATTACAC"
BASE COUNT 112 a 83 c 113 g 183 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BI280645/rev ..
Align seg 1/1 to reverse of: BI280645 from: 1 to: 491

1006 GlnPheAspAspAspPleuSerVal 1014
|||||
475 CAATTGACGATGATGACCTCTCTGTA 449

seq_name: gb_est1:BE169127

seq_documentation_block:
LOCUS BE169127 493 bp mRNA EST 21-JUN-2000
DEFINITION PM3-HT0520-230200-002-h05 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE169127

VERSION BE169127.1 GI:8631848
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 493)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,J.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=gt2=PM3-HT0520-230
200-002-005<3=2000-02-23<4=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 492.
Location/Qualifiers
FEATURES
source
1..493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="HT0520"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 155 a 86 c 106 g 146 t
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x BE169127/rev ..
Align seg 1/1 to reverse of: BE169127 from: 1 to: 493
362 LeuphenisthrPhetYrAspleu1eu 370
|||||
37 CTATTTCATACCTCTCATGACCTTTG 11
seq_name: gb_est1:A1417920
seq_documentation_block:
LOCUS A1417920 505 bp mRNA EST 30-MAR-1999
DEFINITION U955408.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2112687 3',
mRNA sequence.
ACCESSION A1417920
VERSION A1417920.1 GI:4261424
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 503)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds@email.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone-distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrp/image/image.html
Insert length: 649 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 457.
Location/Qualifiers
FEATURES
source
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2112687"
/clone_1lb="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker. Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonesids
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 167 a 89 c 84 g 165 t
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x A1417920 ..
Align seg 1/1 to: A1417920 from: 1 to: 505
700 VALAAGlySerLeuSerLeuVal1eu 708
|||||
205 GTTGCTGGAAGCCTTCTCTGCTTTG 231
seq_name: gb_est1:A1697488
seq_documentation_block:
LOCUS A1697488 506 bp mRNA EST 18-DEC-1999
DEFINITION w614609.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2341048 3',
mRNA sequence.
ACCESSION A1697488
VERSION A1697488.1 GI:4985388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 506)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9abbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/biopr/image/image.html
Insert length: 829 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 455.

FEATURES
source
1. 506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2341048"
/clone_lib="NCI CGAP Lu24"
/tissue_type="Carcinoid"
/lab_host="PH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP-Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clones
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT
167 a 89 c 84 g 166 t

ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x A1697488 ..

Align seg 1/1 to: A1697488 from: 1 to: 506

700 ValAlaGlySerLeuSerLeuValLeu 708
|||||
206 GTTGGCTGGAAGCCTTCTCTGTTTG 232

seq_name: gb_est2:BI280866

seq_documentation_block:
LOCUS BI280866 506 bp mRNA EST 19-JUL-2001
DEFINITION UI-R-DC0-bzp-a-12-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
BI280866
ACCESSION BI280866
VERSION BI280866.1 GI:14930054
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 506)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
MEDLINE
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized seminal vesicles library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-39,
>POLY A#simple-repeat
Seq primer: MJ3 Forward
POLY A=yes

FEATURES
source
1. 506
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DC0-bzp-a-12-0-UI"
/clone_lib="UI-R-DC0"
/dev_stage="ADULT"
/lab_host="PH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-DC0
library is a non-normalized library constructed from rat
seminal vesicle tissue. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-DC0
TAG_TISSUE=seminal vesicles
TAG_SEQ=GTGATTACAC"

BASE COUNT
111 a 84 c 115 g 196 t

ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x BI280866/rev ..

Align seg 1/1 to reverse of: BI280866 from: 1 to: 506

1006 GlnPheAspAspAspAspLeuSerVal 1014
|||||
493 CAATTTCACGATGATGACCTCTGTGTA 467

seq_name: gb_est2:BI280541

seq_documentation_block:
LOCUS BI280541 513 bp mRNA EST 19-JUL-2001
DEFINITION UI-R-DC0-bz1-f-09-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
BI280541
ACCESSION BI280541
VERSION BI280541.1 GI:14929429
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 513)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: mssoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized seminal vesicles library cDNA Library Preparation: M. B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

FEATURES

source

1. .513
Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-DC0-bz1-f-09-0-UI"

/clone_lib="UI-R-DC0"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0 library is a non-normalized library constructed from rat seminal vesicle tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=UI-R-DC0

TAG_TISSUE=seminal vesicles

TAG_SEQ=GTGATTACAC"

BASE COUNT 120 a 120 a 89 c 115 g 189 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BI280541/rev ..

Align seg 1/1 to reverse of: BI280541 from: 1 to: 513

1006 Glnpheaspaspaspaspleuserval 1014

|||||
475 CAATTGACGATGACCTCTCTGTA 449

seq_name: gb_est2:BI280933

seq_documentation_block:

LOCUS BI280933 522 bp mRNA EST 19-JUL-2001
DEFINITION UI-R-DC0-bzo-b-01-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
UI-R-DC0-bzo-b-01-0-UI 3', mRNA sequence.

ACCESSION BI280933
VERSION BI280933.1 GI:14930190

KEYWORDS

SOURCE

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 522)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene

JOURNAL MEDLINE

discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: mssoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized seminal vesicles library cDNA Library Preparation: M. B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

FEATURES

source

1. .522
Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-DC0-bzo-b-01-0-UI"

/clone_lib="UI-R-DC0"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DC0 library is a non-normalized library constructed from rat seminal vesicle tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=UI-R-DC0

TAG_TISSUE=seminal vesicles

TAG_SEQ=GTGATTACAC"

BASE COUNT 120 a 120 a 91 c 117 g 194 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x BI280933/rev ..

Align seg 1/1 to reverse of: BI280933 from: 1 to: 522

1006 Glnpheaspaspaspaspleuserval 1014

|||||
472 CAATTGACGATGACCTCTCTGTA 446

seq_name: gb_est2:BI281386

seq_documentation_block:

LOCUS BI281386 523 bp mRNA EST 19-JUL-2001
DEFINITION UI-R-DC0-bz1-d-11-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
UI-R-DC0-bz1-d-11-0-UI 3', mRNA sequence.

ACCESSION BI281386
VERSION BI281386.1 GI:14931060

KEYWORDS

SOURCE

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 523)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized seminal vesicles library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
 source Location/Qualifiers
 1..523
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DC0-bz0-h-01-0-UI"
 /clone_1lb="UI-R-DC0"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-DC0 library is a non-normalized library constructed from rat seminal vesicle tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_LIB=UI-R-DC0
 TAG_TISSUE=seminal vesicles
 TAG_SEQ=GTGATTACAC"
 BASE COUNT 119 a 93 c 117 g 194 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x BI281386/rev ..

Align seg 1/1 to reverse of: BI281386 from: 1 to: 523

1006 GlnpheaspaspaspplseuSerVal 1014
 |||||
 476 CAATTGACGATGACCTCTCTGTA 450

seq_name: gb_est2:BI280985

seq_documentation_block:
 LOCUS BI280985 527 bp mRNA EST 19-JUL-2001
 DEFINITION UI-R-DC0-bz0-h-01-0-UI.s1 UI-R-DC0 Rattus norvegicus cDNA clone
 ACCESSION UI-R-DC0-bz0-h-01-0-UI 3', mRNA sequence.
 VERSION BI280985.1 GI:14930286
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
 1 (bases 1 to 527)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized seminal vesicles library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
 source Location/Qualifiers
 1..527
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DC0-bz0-h-01-0-UI"
 /clone_1lb="UI-R-DC0"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-DC0 library is a non-normalized library constructed from rat seminal vesicle tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_LIB=UI-R-DC0
 TAG_TISSUE=seminal vesicles
 TAG_SEQ=GTGATTACAC"
 BASE COUNT 120 a 90 c 119 g 198 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x BI280985/rev ..

Align seg 1/1 to reverse of: BI280985 from: 1 to: 527

1006 GlnpheaspaspaspplseuSerVal 1014
 |||||
 472 CAATTGACGATGACCTCTCTGTA 446

seq_name: gb_est1:AM794860

seq_documentation_block:
 LOCUS AM794860 531 bp mRNA EST 16-MAY-2000
 DEFINITION RC6-UM0015-280300-013-H08 UM0015 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM794860
 VERSION AM794860.1 GI:7846834
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 531)
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 COMMENT Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=et2-RC6-UM0015-280300-013-H08&t3=2000-03-28&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 93
 Location/Qualifiers
 1..531
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UM0015"
 /dev_stage="Adult"
 /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 182 a 98 c 84 g 167 t
 ORIGIN
 alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x AM794860 ..
 Align seg 1/1 to: AM794860 from: 1 to: 531
 700 ValAlaGlySerLeuSerLeuValLeu 708
 ||||||||||||||||||||||||||||
 467 GTTGGTGAAGCCTTCTCTGCTTTTG 493
 seq_name: gb_est1:AL509943
 seq_documentation_block:
 LOCUS AL509943 532 bp mRNA EST 04-JAN-2001
 DEFINITION AL509943 Hordeum vulgare Barke developing caryopsis (3'-15.DAP)
 ACCESSION AL509943
 VERSION AL509943.1 GI:12036446
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 532)
 AUTHORS Michael, W., Weschke, W., Pleissner, K.-P. and Graner, A.
 TITLE Unpublished (2000)
 COMMENT Contact: Michael W
 Institute for Plant Genetics and Crop Plant Research
 Corrensstr.3, D-06466 Gatersleben, Germany
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
 Seq primer: M13uni primer for 3'end.
 Location/Qualifiers
 1..535
 /organism="Hordeum vulgare"

AUTHORS Michael, W., Weschke, W., Pleissner, K.-P. and Graner, A.
 TITLE EST sequencing and analysis in barley
 JOURNAL Unpublished (2000)
 COMMENT Contact: Michael W
 Institute for Plant Genetics and Crop Plant Research
 Corrensstr.3, D-06466 Gatersleben, Germany
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
 Seq primer: M13uni primer for 3'end.
 Location/Qualifiers
 1..532
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone_lib="HY03P21u"
 /clone_lib="Hordeum vulgare Barke developing caryopsis (3'-15.DAP)"
 /tissue_type="developing caryopsis (3'-15.DAP)"
 /lab_host="XLDLR"
 /note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2: XhoI; mRNA was made from developing caryopsis (3'-15.DAP) of spring barley variety 'Barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb
 Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'- and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp"
 BASE COUNT 127 a 114 c 120 g 170 t 1 others
 ORIGIN
 alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x AL509943/rev ..
 Align seg 1/1 to reverse of: AL509943 from: 1 to: 532
 451 AlAGlYArGArGleuSerLeuLeu 459
 ||||||||||||||||||||||||||||
 239 GCGGGAAGAGCCTTCATCGCTCTT 213
 seq_name: gb_est1:AL511550
 seq_documentation_block:
 LOCUS AL511550 535 bp mRNA EST 04-JAN-2001
 DEFINITION AL511550 Hordeum vulgare Barke developing caryopsis (3'-15.DAP)
 ACCESSION AL511550
 VERSION AL511550.1 GI:12038053
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 535)
 AUTHORS Michael, W., Weschke, W., Pleissner, K.-P. and Graner, A.
 TITLE EST sequencing and analysis in barley
 JOURNAL Unpublished (2000)
 COMMENT Contact: Michael W
 Institute for Plant Genetics and Crop Plant Research
 Corrensstr.3, D-06466 Gatersleben, Germany
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
 Seq primer: M13uni primer for 3'end.
 Location/Qualifiers
 1..535
 /organism="Hordeum vulgare"

```

/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY08623u"
/clone_lib="Hordeum vulgare Barke developing caryopsis
(3.-15.DAP)"
/tissue_type="developing caryopsis (3.-15.DAP)"
/lab_host="XLOLR"
/note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:
XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
of spring barley variety 'Barke', a high quality malting
variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
(3'-end of cDNA). NOTE: Due to a cloning artefact caused
by the kit, in most cases the EcoRI site is NOT present,
as well as the EcoRI adapter. Average insert size is 1 kb
Sequence trimming: Vector sequences and sequence ends were
trimmed from the 5'-and 3'-end until a 50 bp window
contains less than two ambiguities. The maximum length was
set to 700 bp"
BASE COUNT      134 a      121 c      127 g      152 t      1 others
ORIGIN

```

```

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

```

```

alignment_block:
US-09-697-089-2 x AL511550/rev ..

```

```

Align seg 1/1 to reverse of: AL511550 from: 1 to: 535

```

```

451 AlagIYArgIeuSerserIeu 459
|||||
225 GCGGAGAGAGGCTTTCATCGCTCT 199

```

us-09-697-089-2.oli.go.rng

/SIDS2/gcgdata./geneseq./geneseq./Nm2001..DAT:AA142181 +	8.00	116.13	215.63	509
/SIDS2/gcgdata./geneseq./geneseq./Nm2001..DAT:AAH12617 +	8.00	115.80	224.97	532
/SIDS2/gcgdata./geneseq./geneseq./Nm1999..DAT:AAH28297 -	8.00	115.02	248.78	591
/SIDS2/gcgdata./geneseq./geneseq./Nm2000..DAT:AA95560 +	8.00	114.39	269.70	643
/SIDS2/gcgdata./geneseq./geneseq./Nm2000..DAT:AAFI1778 -	8.00	114.37	270.10	644
/SIDS2/gcgdata./geneseq./geneseq./Nm2001..DAT:AAH42838 -	8.00	114.31	272.51	650

/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:AAH68087 +	8.00	113.83	289.74	693
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH5084 +	8.00	113.16	315.71	758
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH07812 +	8.00	113.11	317.71	763
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH07419 +	8.00	112.42	347.15	837
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH44572 +	8.00	112.26	354.29	855
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH41255 +	8.00	111.43	393.87	955
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH41524 +	8.00	110.54	441.53	1076
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH35424 +	8.00	110.53	442.31	1078
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH45763 +	8.00	110.45	446.63	1088
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH33764 +	8.00	110.16	463.90	1133
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH51511 +	8.00	110.16	463.90	1133
/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:AAH46074 +	8.00	110.03	471.34	1155
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH23354 +	8.00	109.60	497.95	1220

/SID52/gcgata/geneseq/geneseq/NM1996.DAT	AAT29171	8	109.35	514.74	1265.61
/SID52/gcgata/geneseq/geneseq/NM1996.DAT	AAT29172	8	109.35	514.74	1265.61
/SID52/gcgata/geneseq/geneseq/NM2001.DAT	AAD03929	8	109.01	537.74	1322.04
/SID52/gcgata/geneseq/geneseq/NM2000.DAT	AAC76711	8	108.50	579.51	1414.14
/SID52/gcgata/geneseq/geneseq/NM2000.DAT	AAC64484	8	108.02	603.97	1508.1
/SID52/gcgata/geneseq/geneseq/NM1995.DAT	AAO87717	8	108.00	611.52	1512.00
/SID52/gcgata/geneseq/geneseq/NM1995.DAT	AAT28383	8	108.00	611.52	1512.00

/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:ACC32893 +	8.00	108.00	611.90	1513
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21023 +	8.00	107.73	633.57	1555

[illegible]

XX Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
 KW cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
 KW systemic lupus erythematosus; arthritis; neurological disorder; stroke;
 KW Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
 KW aplastic anaemia; myocardial infarction; inflammatory disorder;
 KW Crohn's disease; insulin-dependent diabetes; contact dermatitis;
 KW psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
 KW tuberculosis; ischemic brain injury; hypoxic brain injury; ss;
 KW kidney ischemia; reperfusion injury; acute bacterial meningitis;
 KW excitotoxic brain damage; liver disease.

OS Homo sapiens.

	Key	Location/Qualifiers
FH	CDS	36..3110
FT		

PN WO200130971-A2.

PD 03-MAY-2001.

PF 26-OCT-2000; 2000WO-US29643.

PR 27-OCT-1999; 99US-0161822.

PA (MILL-) MILLENNIUM PHARM INC.

PI Bertin J, Robison KE;

DR WPI; 2001-308628/32.

XX

PT Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -

PS Claim 2; Fig 1; 93pp; English.

The sequence represents a cDNA which encodes the human caspase recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a number of proteins that transmit signals that activate apoptosis and inflammatory pathways in response to stress and other stimuli. Therefore, CARD-12 and its corresponding nucleic acid may be used in treatment and diagnosis of patients suffering from disorders associated with an abnormal level (an increase or a decrease) of apoptotic cell death or abnormal activity of stress-related pathways. The disorders include cancer, viral infections (e.g. caused by poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus erythematosus, arthritis), neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial infarction, stroke), inflammatory and immune system disorders (e.g. Crohn's disease, insulin-dependent diabetes, contact dermatitis, psoriasis, graft rejection), bacterial infections (e.g. tuberculosis, leprosy), ischaemic and hypoxic brain injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial meningitis and liver disease.

Sequence 3133 BP; 903 A; 691 C; 729 G; 810 T; 0 other;

```
alignment_scores:
  quality: 1024.00
  ratio: 1.000
  percent_similarity: 100.000
  percent_identity: 100.000
```

alignment_block:

US-09-697-089-2 x AAS03945

Align seg 1/1 to: AAS03945 from: 1 to: 3133

1	MetAspPheIleIlyAspAspSerArgAlaLeuIleGlnArgMetClyHe	17
36	ATGAATTTTCATAAAGACAAATGCGGAGCCCTTATTCAAAGATGGGAAT	85
17	TThrValIleIlyGsnIleThiAspAspPhePheValTrpAsnValLeuA	34
86	GACGTCTTATAAGCAAAATCACAGATGACCTATTGTTGATGGATGTTCTGA	135
34	snArgIuGluValAsnIleIleCysGlyLysValGluGlnAspAla	50
136	ATCGGAGAAAGTAAACATCATTTGCTGCGAGAGGTGGAGCAGCATGCT	185
51	AlaArgGlyIleIleHisMetIleLeuLysGlySerGlnSerCysAs	67
186	GCTAAGGAGATCATTCACATGATTTTGAAGAAAGGTTTCAGAGCTCTGA	235
nLeuPheLysSerLeuLysGlnTrpAsnTrpProLeuPheGlnAspL	84	
236	CTCTCTTTTAATCCCTTAAGGAGGGAACATGCTCTATTTCAGAGACT	285
84	euAsnGlyGlnSerLeuPheHisGlnThrSerGluLysPheAspAsp	100
286	TGAATGGACAAGCTTTTTCATCAGACATCAGAAGGACACTTGGACGAT	335
101	LeuAlaGlnAspLeuLysAspLeuTrpHisThrProSerPheLeuAsnPh	117
336	TTGGCTCAGGATTTTAAGAGACTTGATACCATCCCCATCTTTCTGAACTT	385
117	eTyrProLeuGlyLysAspIleAspIleIlePheAsnLeuLysSerThrP	134
386	TTATCCCTTGGTGAAGATATTGACATTATTTTAACTTGAAAGGACCT	435
134	heThrGluProValLeuTrpArgLysAspGlnHisHisArgValGlu	150
436	TCACAGAACTGCTCTGAGAGAGAGCAACACATCAGCGCGTGGAG	485
151	GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI	167
486	CAGCTGACCCCTGAAGGCGCTCTGAGGCTCTTCAGAGCCCTGCTCAT	535
167	eGluGlyGlnSerGlyLysGlyLysSerThrLeuGlnArgIleAlaM	184
536	TGAAGGGAAATCTGGCAAAAGCAAGTCCACTGCTGACAGCGCATTTGCCA	585
184	eLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheVal	200
586	TGCTCTGGGGGCTCGGAAAGTGCMAAGGCTGACCAAGTTCAAAATTCGTC	635
201	PhePheLeuArgLeuSerArgAlaGlnGlyLeuPheGlnThrLeuCy	217
636	TTCTCTCCCTCCCTCAGAGGGCCAGGGGTGACATTTTGAACCCCTGTG	685
217	sAspGlnLeuAspIleProGlyThrIleArgLysGlnThrPheMetA	234
886	TGATCAACTCTCGGATTAATCTGGGCAACAATCAGAAAGACATTCATGG	735
234	IaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuAspGlyTrp	250
736	CCATGCTCTGTAAGTGGGGCAGAGAGTCTTTCTCTTTGATGGCTAC	785
251	AsnGlnPheLysProGlnAsnCysProGlnIleGlnAlaLeuIleLysG	267
786	AATGAAATTCAGAGCCGAGAACTGCCAGAAATCGAAAGCCGTATAAAGA	835
267	uAsnHisArgPheLysAsnMetValIleValThrThrThrTrpGluCyl	284
836	AAACACACCGCTTCAAGAACATGGGTATGCTGCACCACTACACGTGATGCC	885
284	euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet	300
886	TGAGGCAATATGCGGAGTTTGGTGGCCCTGACTGCTGAGGTGGGGGATATG	935
301	ThrGluAspSerAlaGlnAlaLeuIleArgGluValIleuIleLysGlu	317

```

936 ACAGAAAGACAGCCGACGCTCTCATCGAGAGAGCTGATCAAGAGACT 985
317 uAlaGlUGlUleuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
986 TGCtGAAGGCTTGTGCTCCAAATTCAGAAATCCAGAGTGTGAGGAATC 1035
334 eumEtLysThrProLeuPheValIleThrCysAlaIleGlnMeGly 350
1036 TCATGAAGACCCCTCTCTTGTGGTCAATCACTTGGCATCCAGATGGGT 1085
351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
1086 GAAAGTAGTTCACACTCTCACACACAACAACGCGTGTCCATACCTCTTA 1135
367 rAsPLeuLeuIleGlnLysAsnLysHisLysHisLysGlyAlaAlaAs 384
1136 TGATCTGTGTACAGAAAAACAACAACATAAAGGTGTGGCTGCA 1185
384 eRAsPHeiLeArgSerLeuAspHisCysGlyAspLeuAlaLeuGlnGly 400
1186 GTGACTTCATTCGAGACCTGGACCACTGTGGAGCCTTAGCTGGAGGGT 1235
401 ValPheSerHisLysPheAspPheGlnLeuGlnAspValSerSerValAs 417
1236 GTGTTCTCCCAAGATTGATTTCGAAGTGCAGAGATGTCTCAGCGTGA 1285
417 nGUAspValLeuLeuThrThrGlyLeuLeuCysLysTyrrThrAlaGln 434
1286 TGAGGATGTCTGCTGACAACTGGGCTCTCTTAAATATACACTCAAA 1335
434 rGpHeLysProLysTyrrLysPhePheHisLysSerPheGlnGluTyrrThr 450
1336 GGTTCAAAGCCAAAGTATATAATTTCTTCACAAGTATTCAGAGATACCA 1385
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGlnGluVal 467
1386 GCAGGACGAAAGACTGACAGATTATGTGACGTCTCATGAGCCAGAGAGT 1435
467 lThLysGlyAsnGlyTyrrLeuGlnLysMetValSerIleSerAspIleT 484
1436 GACCACAGGGGAATGTTACTTGCAAAATGTTCCATTTCGAGACATTA 1485
484 hrSerThrTyrrSerSerLeuLeuArgTyrrThrCysGlySerSerValGlu 500
1486 CATCACTTATAGCAGCCTGCTCCGGTACACCTGTGGGTCTCTGTGAA 1535
501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrrGlnHisGlyCy 517
1536 GCCACACAGGCGTGTATGAACACCTGCGACAGCTGATCAACACGCGCTG 1585
517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuThrArgGlnGlnSerL 534
1586 CTTTCTGGAGCTTTCATCGCCAAAGAGCGCTCTGTGGAGACAGGAATCTT 1635
534 euGlnSerValLysAsnThrThrGlnGlnGlnIleLeuLysAlaIleAsn 550
1636 TGCAAAGGTGAAAAACACACACCTGACGAGAAGAAATTCGAAAGCCATAAC 1685
551 lLeAsnSerPheValGluCysGlylLeHisLeuTyrrGlnGlnSerThrSe 567
1686 ATCAATTCCTTTGTAGAGTGTGCGATCCATTTATATCAAGAGATACATC 1735
567 rLySerAlaLeuSerGlnGlnGlnGlnGlnAlaPhePheGlnGlnLysSerL 584
1736 CAAATTCAGCCCTGAGCCAAAGAAATTTGAAGCTTTCTTCAAAGTAAACCT 1785
584 euTyrrlLeAsnSerGlyAsnIleProAspTyrrLeuPheAspPheGln 600
1786 TATATATCAACTCAGGGACATCCCGATTTACTTATTTGACTTCTTTGAA 1835
601 HisLeuProAsnCysAlaSerAlaLeuAspHeiLeLysLeuAspPheTy 617
|||||
1836 CATTTGCCCAATGTGCAAGTGTCTGACTTCATTAACATGACTTTTA 1885
617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyTy 634
1886 TGGGGAGACTATGGCTTCATGGGAAAAAGCTGCAGAAAGACAGAGTGGAA 1935
634 lHisMetGluGluAlaProGluThrTyrrIleProSerArgAlaValSer 650
1936 TCCACATGGAAGAGAGCCCGCAAAACCTACATTCACAGAGGCGCTGTATCT 1985
651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGlnValAlaThrLeuAr 667
1986 TTGTTCTTCAACTGGAAGACAGCAATTCAGACCTGTGGAGGTACACATCCG 2035
667 gAsPPheserLysLeuAsnLysGlnAspIleThrTyrrLeuGlnLysIleP 684
2036 GGATTTCAAGCAAGTTGAATAGCAAGATATCAATATCTGGGGAAATAT 2085
684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
2086 TCAGCTGTGCCACAAAGCTCTAGGCTGCCAATTAAGAGATGTGCTGTGTG 2135
701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrrSerLe 717
2136 GCTGGAAGCCTCAGATTGTGCTCAGACACCTGTGAAGAACATTTATCTCT 2185
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
2186 CATGCTGGAAGCCAGTCCCTCCACCATPAGAAGATGAGAGGCACTACAT 2235
734 eValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
2236 CTGTAAACAACCTGAAAACCTTGAGTATTCATGACCTACAGAACTCAACGG 2285
751 LeuProGlyGlyLeuThrAspSerLeuGlyLysLeuLysAsnLeuThrLy 767
2286 CTGCCGGGTGTGTGACTGACGACCTGGGTACTTGAAGAACCTTACAAA 2335
767 sLeuIleMetLysPAsnIleLysMetAsnGlnGluAspAlaIleLysLeuA 784
2336 GCTCATATATGATTAACATTAAGATGAATGAAGAAAGATGCTATAAATAG 2385
784 lAsnGlyLysLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
2386 CTGAAGGCCCTGAAAAACCTGAAGAAGATGTGTTATTTCAATTTACACCC 2435
801 LeuSerAspIleGlyGluGlyMetAspTyrrIleValLysSerLeuSerSe 817
2436 TTGCTGTACATTTGAGAGAGGCAATGATTTACATAGTCAAGTCTGTCAAG 2485
817 rGluProCysAspLeuGlnGlnIleGlnLeuValSerCysCysLeuSerA 834
2486 TGAACCCGTGTACCTTGAAGAAATTCATTAATGTCTCCGTGCTGTGTG 2535
834 lAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
2536 CAATATGACGTAAAAATCTTACCTCAGAAATCTTCACAAATTTGGTCAAACTG 2585
851 SerIleLeuAspLeuSerGlyAsnTyrrLeuGlnLysAspGlyAsnGluAl 867
2586 AGCATTTCTTGAATTATCAGAAAAATTTACCTGGAAAAAGATGGAATGANA 2635
867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAla 884
2636 TCTTCATGATACGTATCGACAGATGAACGTGCTGAACAGCTCACCCAC 2685
884 eumEtLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
2686 TGATGCTGCCCTGGGGCTGTGACGTGCAAGGACACCTGAGAGGCTGTGTG 2735
901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
2736 AAACATTTGAGAGAGTCCCAACAACGTCTCAAGCTTGCGTTGGAATAACTG 2785

```

```

917  pArgLeuThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLysa 934
|||||
2786  GAGACTCACAGATACAGAGATTGAAATTTAGTGCATTTTGGAAAGC 2835
934  snProLeuYsaSnPheGlnGlnLeuAsnLeuAlaGlySnaArgValSer 950
|||||
2836  ACCCTCTGAAAAAAGCTTCCAGCAGTTGAAATTTGGCGGAAATCGTGAGC 2885
951  SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuYsaGlnle 967
|||||
2886  ACTGATGATGATGGCTTCCCTTCATGGGTGATTTGACAATCTTACGAAT 2935
967  uValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeuV 984
|||||
2936  ACTGTTTGTGACTTACTAGTAAAGAAATTTCTACCTGATCCAGCATAG 2985
984  aArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
|||||
2986  TCAGAAAGAACTTAGCCAGAGTGTATCCAACTTAACTTTCTGCAGAGAGCT 3035
1001  ArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIleThrG 1017
|||||
3036  AGCGTTGTGGGTGGCAATTTGATGATGATGATGATGATGATGATGATGAT 3085
1017  yAlaPheLysLeuValThrAla 1024
|||||
3086  TGCCTTTTAAACTAGTAACTGCT 3107

seq_name: /SIDS2/gcdata/geneseq/geneseq/M2001.DAT:AA503946
seq_documentation_block:
ID   AA503946 standard; DNA; 3615 BP.
XX
XX   AA503946;
AC
AC   12-SEP-2001 (first entry)
DE
DE   Human caspase recruitment domain 12 (CARD-12) genomic DNA.
XX
XX   Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW   cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW   systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
KW   aplastic anaemia; myocardial infarction; inflammatory disorder;
KW   Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW   psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW   tuberculosis; ischaemic brain injury; hypoxic brain injury; ds;
KW   kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KW   excitotoxic brain damage; liver disease.
XX
XX   Homo sapiens.
OS
OS   Key
FH   Location/Qualifiers
FT   CDS
FT       1..3615
FT       /*tag= a
FT       /product= "Human CARD-12"
XX
XX   WO200130971-A2.
XX
XX   03-MAY-2001.
XX
XX   26-OCT-2000; 2000WO-US29643.
XX
XX   27-OCT-1999; 99US-0161822.
XX
XX   (MILL-) MILLENNIUM PHARM INC.
XX
XX   Bertin J, Robison KE;
PI
PI   WPL: 2001-308628/32.
DR
DR   P-PSDB; AAU02881.
XX

```

```

PT Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -
XX
XX   Disclosure: Fig 2; 93pp; English.
PS
PS   The sequence represents a genomic DNA which encodes the human caspase
CC   recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
CC   number of proteins that transmit signals that activate apoptosis and
CC   inflammatory pathways in response to stress and other stimuli. Therefore,
CC   CARD-12 and its corresponding nucleic acid may be used in treatment and
CC   diagnosis of patients suffering from disorders associated with an
CC   abnormal level (an increase or a decrease) of apoptotic cell death or
CC   abnormal activity of stress-related pathways. The disorders include
CC   cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
CC   autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC   neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC   sclerosis), hematologic diseases (e.g. aplastic anaemia, myocardial
CC   infarction, stroke), inflammatory and immune system disorders (e.g.
CC   Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC   psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC   lepromatous leprosy), ischaemic and hypoxic brain injury, kidney
CC   ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC   meningitis and liver disease.
XX
XX   Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;
SQ
SQ

```

Alignment scores:

Alignment	Quality	Ratio	Length	Gaps
US-09-697-089-2 x AA503946	827.00	0.892	928	0
Percent Similarity:	99.892	Percent Identity:	99.892	

Alignment block:

```

US-09-697-089-2 x AA503946
..
Align seg 1/1 to: AA503946 from: 1 to: 3615
89  LeuPheHisGlnThrSerGluGlyAspLeuAspAspLeuAlaGlnAspLe 105
|||||
736  CTTTTCATCACACATCAGAGAGACTTGGAGATTTGGCTCAGAGATT 785
|||||
105  uLysAspLeuTrpHisThrProSerPheLeuAsnPheTrpProLeuGlyG 122
|||||
786  AAAGAGCTGTGACATACCCCATCTTTCTGAACCTTATTCCTGGTG 835
|||||
122  LuAspLeuSplIleIlePheAsnLeuLysSerThrPheThrGluProVal 138
|||||
836  AAGATATTGACATTTATTTTAACTTGAAAGAACCCCTTCACAGAACCTG 885
|||||
139  LeuTrpArgLysAspGlnHisThrHisArgValGlnGlnLeuThrLeuAs 155
|||||
886  CTGTGAGAGAGAGACCAACCATCACCGCTGGAGACCTACCCCTGAA 935
|||||
155  nGlyLeuLeuGlnAlaLeuGlnSerProCysIleIleGlnGlyGluSerG 172
|||||
936  TGGCCTCTGCAAGGCTCTTCAGAGCCCTGCATCATTTGAAGGGAATCG 985
|||||
172  LyLysGlyLysSerThrLeuLeuGlnArgIleAlaMetLeuTrpGlySer 188
|||||
986  GCAAAGCAAGTCACACTGCTGCAGCCGAATTGCCATGCTCTGGGGCTTC 1035
|||||
189  GlyLysCysLysAlaLeuThrLysPheLysPheValPhePheLeuArgLe 205
|||||
1036  GGAAAGTGCAGAGCTCTGACCAAGTTCAATTCGTCCTTCTCCGCTGT 1085
|||||
205  uSerArgAlaGlnGlyGlyLeuPheGluThrLeuCysAspGlnLeuLeu 222
|||||
1086  CAGCAGGGCCAGAGGTGAGACTTTTGAAMCCCTCTGTGATCAACCTCG 1135
|||||
222  SplIleProGlyThrIleArgLysGlnThrPheMetAlaMetLeuLeuLys 238
|||||

```

1136 ATATACCTGGCACATTCAGAGACAGACATTCATGCGCATCTGCTGAG 1185
239 LeuArgGlnArgValLeuPheLeuLeuAspGlyTyrAsnGluPheLysPro 255
1186 CTGCGGAGAGGGTCTTTCTCTTCTTGATGCGCTACAAATGATTCAGAGCC 1235
255 oglnAsnCysProGluIleGluAlaLeuIleLysGluAsnHisArgPheL 272
1236 CCAGAACTGCCAGAAATCGAAGCCCTGATAAAGAAACACCGGCTTCA 1285
272 yAsnMetValIleValThrThrThrGluCysLeuArgHisIleArg 288
1286 AGAACATGGTCATCTCCACCACTGAGTGCCTGAGGACATATCGG 1335
289 GluPheGlyAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAl 305
1336 CAGTTTGTCCTGCTGACTGAGGTGGGATATGACAGAGACACAGCG 1385
305 acIlnAlaLeuIleArgGluValLeuIleLysGluLeuAlaGluGlyLeuL 322
1386 CCAGGCTCTCATCCGAGAAAGTGGTGATCAAGAGAGCTGCTGAAGGCTTGT 1435
322 euLeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 338
1436 TGCTCCAAATTCAGAAATCCAGGTGCTTGAGGAATCTCATGAAGACCCCT 1485
339 LeuPheValValIleThrCysAlaIleGluMetGlyLysSerGluPheHis 355
1486 CTGTTTGCTGATCATCTTGTCATTCAGATGGGTAAAGTGAGTTCCA 1535
355 sSerHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleG 372
1536 CTCTCACACAAACAACAGCGCTGTTCCATCTCTTATGATCTGTGATAC 1585
372 IlnLysAsnLysHisLysHisLysGlyValAlaAspAspPheIleArg 388
1586 AGAAAAACAACAACAACATAAAGCTGCTGCAAGTCACTTCATTCGCG 1635
389 SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisLys 405
1636 AGCCTGACACCTGTGGAGACCTAGCTCTGGAGGGTGTGTTCTCCACAA 1685
405 sPheAspPheGluLeuGlnAspValSerSerValAsnGluAspValLeuL 422
1686 GTTGGATTTGCAACTGCGAGATGTGTCCACGCGAATGAGAGATGCTCTGC 1735
422 euThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPheLysProLys 438
1736 TGACAACTGGGCTCTCTCTTAATATACAGCTCAAGGTTCAAGCCAAAG 1785
439 TyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGlyArgArgLe 455
1786 TATTAATCTTTTCACAACTTCAGAGATACACAGACAGCAAGACT 1835
455 uSerSerLeuLeuThrSerHisGluProGluGluValThrLysGlyAsnG 472
1836 CACGAGTTTATGACGTCATGACGCCAGAGAGGTATCCACCAAGGGGATG 1885
472 IlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerThrTyrSer 488
1886 GTTACTTGGCAGAAATGCTTCCATTTCCGACATTCACATCCACTTATAGC 1935
489 SerLeuLeuArgTyrThrCysGlySerSerValGluAlaThrArgAlaVal 505
1936 AGCCTGCTCGGTACACTGTGGGTCACTGTGGAAACCCACAGGGGTGT 1985
505 IlnetLysHisLeuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLeus 522
1986 TATGAAGACACTCGCAGACAGTATATCAACAGGCTGCTCTCGGAGATT 2035
522 erIleAlaLysArgProLeuTTPArgGlnGluSerLeuGlnSerValLys 538
2036 CCATCGCCAAAGAGGCTCTCTGAGACAGGAATCTTTGCAAAAGTGTA 2085

539 AsnThrThrGluGlnGluIleLeuLysAlaIleAsnIleAsnSerPheVal 555
2086 AACACCACTGAGCAAGAAATTTCTGAAAGCCATTAACATCAATTCCTTTGT 2135
555 IgluCysGlyLysIleHisLeuTyrGlnGluSerThrSerLysSerAlaLeus 572
2136 AGAGTGTGGCATCTTATATATCAAGAGATCAATCCAAATCAGCCCTGA 2185
572 erGlnGluPheGluAlaPhePheGlnGlyLysSerLeuTyrIleAsnSer 588
2186 GCCAAGATTTGAAGCTTCTTCACAGTAAACCTTATATATCACTCA 2235
589 GlyAsnIleProAspTyrLeuPheAspPheGlnHisIleLeuProAsnCy 605
2236 GGGAACTCCCGGATTTACTTATTTGACTTCTTTAACAATTTGCCCAATTG 2285
605 sAlaSerAlaLeuAspPheIleLysLeuAspPheTyrGlyAlaMetAla 622
2286 TGCAAGTGGCCCTGGACCTTCAATTAACCTGGAAGCTTTATGGGGAGCTATGG 2335
622 IAserTTPGluLysAlaAlaGluAspThrGlyGlyIleHisMetGluGlu 638
2336 CTTCATGGGAAAGGCTGCAGAAAGACACAGGTGGAATCCACATGGAAGAG 2385
639 AlaProGluThrTyrIleProSerArgAlaValSerLeuPhePheAsnTr 655
2386 GCCCAGAAACCTACATTCCTCCAGCAGGCGTGTATCTTTGTTCTTCAACTG 2435
655 pLysGlnGluPheArgThrLeuGlnGluValThrLeuArgAspPheSerLysL 672
2436 GAAGCAGGAATTCAGSACTCTGGAGGTACACACTCGGGATTTTCAGCAGT 2485
672 euAsnLysGlnAspIleThrTyrLeuGlnLysIlePheSerSerAlaThr 688
2486 TGAATTAAGCAAGATATCAGATATCTGGGAAATATTCACCTCGCCACA 2535
689 SerLeuArgLeuGlnIleLysArgCysAlaGlyValAlaGlySerLeuSe 705
2536 AGCCTCAGGCTGCCAAATTAAGAGATGTGCTGTGCTGCGTGAAGCCCTCAG 2585
705 rLeuValLeuSerThrCysLysAsnIleTyrSerLeuMetValGluAlaAs 722
2586 TTTGGTCTCTCAGCACCTGTGAAGAACATTTATCTCTCATGTGTGAAGCCA 2635
722 erProLeuThrIleGluAspGluArgHisIleThrSerValThrAsnLeu 738
2636 GTCCCTCCACCATATGAAGATGAGAGGCACATCATCTGTAAACAAACCTG 2685
739 LysThrLeuSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLe 755
2686 AAAACCTTGAGTATTCATGACCTACAGATCAACAGAGGCTGGCGGTGTCT 2735
755 urThrAspSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAl 772
2736 GACTGACAGCTTGGGTAACTTGAAGAACCTTACAAACCTCATATATGATA 2785
772 snIleLysMetAsnGluGluAspAlaIleLysLeuAlaGluGlyLeuLys 788
2786 ACATTAAGATGAATGAAGAGATGCTTAAACACAGCTGAAGGCTGMAA 2835
789 AsnLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleG 805
2836 AACCTGAAGAGATGTGTTATTTATTTATTAACCCACTTGTGTGACATTTGG 2885
805 yGluGlyMetAspTyrIleValLysSerLeuSerSerGluProCysAspL 822
2886 AGAGGGAATGATTAATCATAGTCTCTCAAGTGAAGAACCTGTGTAC 2935
822 euGluGluIleGlnLeuValSerCysCysLeuSerAlaAsnAlaValLys 838
2936 TTGAAGAAATTCATTAATAGTCTCTGCTGCTGTGCTGCAAAATGACAGTGAA 2985

```

839 11leuA1aGlnAsnLeuHisAsnLeuValLysLeuSer11leuAspLe 855
|||||
2986 ATCTAGACTCAGAAATCTTCACAAATTTGTCACAACTGACATCTTGTATTT 3035
|||||
855 uSerLysnTYrLeuGluLysAspG1LysnGluAlaLeuHisGluLeu1 872
|||||
3036 ATCAGAAAAATTCCTGGAAAAAGAAATGGAATGAGCTCTTCATGAACCTGA 3085
|||||
872 1eAspArgMetAsnValLeuGluGlnLeuThrAlaLeuMetLeuProTPr 888
|||||
3086 TCGACAGATGAGACGTGCTAGAACACACTCACCGCAGCTGATGCTGCTTGG 3135
|||||
889 GlyCysAspValGlnGlySerLeuSerSerLeuLeuLysHisLeuGlu1 905
|||||
3136 GCGCTGACGTGCAAGCGCAGCTGAGCACCTGTTGAAACATTTGGAGAGA 3185
|||||
905 uValProGlnLeuValLysLeuGlyLeuLysAsnTPArGLeuThAspT 922
|||||
3186 GGTCCCAACACTGTCAGACCTGGGTTGAAAAACTGGAGACTCACAGATA 3235
|||||
922 hGlu11eArG11eLeuG11eAlaPhePheG11eLysAsnProLeuLysAsn 938
|||||
3236 CAGAGATTGAAATTTAGTGCATTTTGGAAAGAACCTCTGAAAAAC 3285
|||||
939 PheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAspGlyTTPLe 955
|||||
3286 TTCACGACAGTGAATTTGCGCGGAAATCCTGTGACAGCATGATGATGCGCT 3335
|||||
955 u1AlaPheMetGlyValPheGluAsnLeuLysGlnLeuValPhePheAsp 972
|||||
3336 TCCCTTCATGGGCTGATTTGAGAAATCTTAAGCAATTTAGTGTTTTGGACT 3385
|||||
972 hSeSerThrLysGlnPheLeuProAspProAlaLeuValArgLysLeuSer 988
|||||
3386 TTAGTACTAAAGAAATTTCTACCTGATCCAGCATTTAGTCAGAAACTTACG 3435
|||||
988 GlnValLeuSerLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTr 1005
|||||
3436 CAAGGTGTTCACAGTTAACTTTCTGCAAGAAAGCTAGGCTTGTGGGTG 3485
|||||
1005 pGlnPheAspAspAspAspLeuSerVal11eThr 1016
|||||
3486 GCATTTTGAATGATGATCTCAGTGTATTATCA 3519
|||||
seq_name: /SIDS2/gcdata/geneseq/geneseq/NA2001.DAT:AAH98254
seq_documentation_block:
ID AAH98254 standard; cDNA; 3545 BP.
AC AAH98254;
XX
XX 12-OCT-2001 (first entry)
DT
XX
DE Murine EST-derived coding sequence SEQ ID NO: 111.
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Mus musculus.
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02687.
XX
XX 25-JAN-2000; 2000US-0491404.
XX 17-JUL-2000; 2000US-0617746.
XX 03-AUG-2000; 2000US-0631451.
XX 15-SEP-2000; 2000US-0663870.
XX

```

```

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WP1: 2001-476164/51.
DR P-PSDB: AAM23595.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
PS Claim 1; Page 250-251; 1275pp; English.
XX
XX The present invention provides the protein and coding sequences of novel
XX proteins from a variety of organisms, including human, dog, cat, horse,
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX from the organism of interest. They can be used in diagnostics,
XX forensics, gene mapping, identification of mutations, to assess
XX biodiversity and for nutritional purposes. The present sequence is a cDNA
XX of the invention.
XX
SQ Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other:

alignment_scores:
Quality: 721.00 Length: 1024
Ratio: 0.706 Gaps: 0
Percent Similarity: 99.707 Percent Identity: 99.707

alignment_block:
US-09-697-089-2 x AAH98254 ..
Align seg 1/1 to: AAH98254 from: 1 to: 3545

1 MetAsnPhe11eLysAspAsnSerArgAlaLeu11eGlnArgMetClyme 17
|||||
232 ATGCAATTTCAATAAGGACAAATAGCCGACCTTATTCAAAGATGGGAAT 281
|||||
17 tThrVal11eLysGln11eThrAspAspLeuPheVal1TPAsnValLeuA 34
|||||
282 GACTGTATTAAGCANAATCACAGATACCTATTGTATGGAATGTTCTGA 331
|||||
34 snArgGluGluValAsn11e11eCysGlyLysValGluGlnAspAla 50
|||||
332 ATGCGCAAGAGTAACATCATTTGCTGCGAGAGGTGAGCAGATGCT 381
|||||
51 AlaArgGly11e11eHisMet11eLysLysGlySerGluSerCysAs 67
|||||
382 GCTAGAGGATCATTCACATGATTTGAAAAAGGTTTCAGAGTCCGTGA 431
|||||
67 nLeuPheLeuLysSerLeuLysGluTPAsnTYrProLeuPheGlnAspL 84
|||||
432 CCTCTTCTTAATCCCTTAAGAGTGAACATACTATCTTATTCAGAGACT 481
|||||
84 euAsnGlyClnSerLeuPheHisGlnThrSerGluGluAspLeuAspAs 100
|||||
482 TGAATGGACAAGTCTTTTTCATCAGACATCAGAAAGAGACTTGGACGAT 531
|||||
101 LeuAlaGlnAspLeuLysAspLeuTYrHisThrProSerPheLeuAsnPh 117
|||||
532 TTGGCTCAGAGATTAAAGACTTGTACCATACCCCTCTTTCTGAAACTT 581
|||||
117 eTYrProLeuGluGluAsp11eAsp11eLeuPheAsnLeuLysSerThp 134
|||||
582 TTATCCCTTGGTGAAGATATTGACATTAATTTTAATCTGAAAAAGCACT 631
|||||
134 hethGluProValLeuThrArgLysAspGlnHisHisArgValGlu 150
|||||
632 TCACACAACTGTCCTGTGAGGAAGACACACATCACCGCGGTGGAG 681
|||||
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCys11e11 167
|||||

```


682 CACTGACCTGAATGGCTCTCTGACAGCTCTTCAGAGACCCCTGCATCAT 731
 167 egluglyGlySerGlyLysGlyLysSerThrLeuGlnArgLLeuLam 184
 732 TGAAGGGGAATCTGGCAAGCAAGTCCACTCTGCTGACGGCAATTCACA 781
 184 etLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
 782 TGCCTGGGGCTCCGGAAAGTGGACAGGCTGTACCAAGTTCAAATTCGTG 831
 832 TTTCTTCCCTCCGTCTACAGAGGGCCAGGCTGAGACTTTTGAACCCCTGTG 881
 217 SASPGlnLeuLeuAspLLeuProGlyThrLLeuArgLysGlnThrPheMet 234
 882 TGATCAACTCTGTGATATACCTGGACAAATCAGAACAGCAATCATATG 931
 234 LameLLeuLysLLeuArgLInArgValLeuPheLeuLeuAspGlyTyr 250
 932 CCATGCTGCTGAAGCTGCGGACAGAGGTTCTTTCTTCTTGATGGCTAC 981
 251 AsnGlnPheLysProGlnAsnCysProGlnLLeuGlnAlaLeuLLeuLysG 267
 982 AATGAATTCAAAGCCCCAGAACTGCCCCAGAAATCGAACCCCTGATTAAGA 1031
 267 uAsnHisArgPheLysAsnMetValLLeuValThrThrThrGlyLysL 284
 1032 AAACCCAGGCTTCAGAAACATGTGTCACTCCACACATCACATGACATGCC 1081
 284 euArgHisLLeuArgLInPheGlyAlaLeuThrAlaGlnValGlyAspMet 300
 1082 TGAAGGCATATCGGCACTTTGGTCCCTGACTGCTGAGGTGGGGGATATG 1131
 301 ThrGlnAspSerAlaGlnAlaLeuLLeuArgLValLeuLLeuLysG 317
 1132 ACAGAAAGACAGCCCCAGGCTCATCTCGAGAAAGTCTGATCAAGAGAGCT 1181
 317 uAlaGlnGlyLLeuLeuGlnLInLeuLysSerArgCysLeuArgAsnL 334
 1182 TGGTGAAGGCTGTGTCTCAAAATTCAGAAATCCAGGTGCTTGAGGAAATC 1231
 334 euMetLysThrProLeuPheValValLLeuThrCysAlaLLeuGlnMetGly 350
 1232 TCATGAAGACCCCTCTCTTGTGTGTCATCTGTGTGCAATCCAGATGGGT 1281
 351 GluSerGlnPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
 1282 GAAAGTGAGTTCACCTCTCACACAAACAACGCTGTTCATACCTCTTA 1331
 367 rAspLeuLeuLLeuGlnLysAsnLysHisLysHisLysGlyValAlaLAs 384
 1332 TGAATCTGTGTATACAGAAAAACAACAACAATTAAGGTGTGCTGCCAA 1381
 384 eArgSPheLLeuArgSerLeuAspHisCysGlyAspLeuAlaLeuGlnGly 400
 1382 GTGACTTCATTCGGAGGCTGGACACACTGTGATACCTAGCTGTGGAGGT 1431
 401 ValPheSerHisLysPheAspPheGlnLeuGlnAspValSerSerValAs 417
 1432 GTGTCTCCCAACAAGTTTCATTCGAACCTCAGAGATGTGTCCAGCGTGA 1481
 417 ngLAspValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGln 434
 1482 TGAAGATGTCTCTGACAACTGGGCTCTCTGTAATATATACAGCTCAA 1531
 434 rgpPheLysProLysTyrLysPhePheHisLysSerPheGlnGlnLysTyrThr 450
 1532 GGTTCAGACCAAGATATAATTTCTTCACAGATCATTCACAGAGTACACA 1581
 451 AlaGlyArgArgLLeuSerSerLeuLeuThrSerHisGlnProGlnGlnVal 467
 1582 GCAGGAGCAAGACTCAGACGTTTATTGACGTCTCATATGAGCCAGAGAGGT 1631

467 LThrLysGlyAsnGlyTyrLeuGlnLysMetValSerLLeuSerAspLLeu 484
 1632 GACCAAGGGGAATGTACTTGCAGAAAAATGTTTCATTTCCATTTGGACATTA 1681
 484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGln 500
 1682 CATCACTTATATAGCAGGCTGCTCGGTATACACTGTGGGTATCATCTGTGGA 1731
 501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
 1732 GCCACACAGGCTGTATGAAGACACCTCGACAGACTGATATCAACACGCGCTG 1781
 517 sLeuLeuGlyLeuSerLLeuAlaLysArgProLeuThrPargGlnGluSerL 534
 1782 CTTCTCTGGACTTTCATGCCCAAGAGGCTCTCTGTGAGACAGAGAACTTT 1831
 534 euGlnSerValLysAsnThrThrGlnGlnGlnLLeuLysAlaLLeuAsn 550
 1832 TGCAAACTGTGAAAAACACCACTGACAGAAATTTGMAAGGCATTAAC 1881
 551 LLeuAsnSerPheValGlyCysGlyLLeuHisLeuTyrGlnGluSerThrSe 567
 1882 ATCAATTCCTTTGTAGAGTGTGGCATCTATTATATCAAGAGATCATATC 1931
 567 rLysSerAlaLeuSerGlnGlnPheGlnValAlaPhePheGlnGlyLysSerL 584
 1932 CAAATTCAGCCCTGAGCCAAAGAAATTTGAAGCTTCTTTCAAGATTAAGCT 1981
 584 euTyrLLeuAsnSerGlyAsnLLeuProAspTyrLeuPheAspPheGln 600
 1982 TATATATCAACTCGAGGACATCCCGCATTAATTGACTTCTTTGAA 2031
 2032 CATTTGCCCAATTTGTGCAGATGCTGTGACTTCATTAACGTGGGCTTTTA 2081
 617 rGlyGlyAlaMetAlaSerTrpGlnLysAlaAlaGlnAspThrGlyLys 634
 2082 TGGGGAGCTATGGCTTCATGGGAAAAAGGCTCGCAAGAACACAGGTGGA 2131
 634 LLeuMetGlnGlnAlaProGlnLLeuThrTyrLLeuProSerArgAlaValSer 650
 2132 TCACATGAGAAAGAGGCCCAAGAAACCTACATTCACACAGGCGGTATCT 2181
 651 LeuPhePheAsnTrpLysGlnGlnPheArgThrLeuGlnValThrLeuArg 667
 2182 TTTGTTCTTCAACTGGAAGCAGGAATTCAGGACTCTGGAGGTTCACACTCCG 2231
 667 gAspPheSerLysLeuAsnLysGlnAspLLeuThrTyrLeuGlyLysLLeu 684
 2232 GGATTTTCAGCAAGTTGATTAAGCAAGATATCAGATATCTGGGAAAAATAT 2281
 684 hSerSerAlaThrSerLeuArgLeuGlnLLeuLysArgCysAlaGlyVal 700
 2282 TCAGCTGTGCACAGAGCTCAGGCTGCAATTAAGAAATATGTGCTGTG 2331
 701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnLLeuTyrSerLe 717
 2332 GCTGGAAGCTCTAGTTGTGCTCTCAGACACCTGTAAAGAACTTTATTTCTCT 2381
 717 uMetValGlnAlaSerProLeuThrLLeuAspGlnArgHisLLeuThr 734
 2382 CATGCTGGAAGCCAGTCCCTCCATATAGAAAGATGAAGGACACATCCAT 2431
 734 eValThrAsnLeuLysThrLeuSerLLeuHisAspLeuGlnAsnGlnArg 750
 2432 CTGTAACAAACCTGMAAACCTTAGATTTATGATGACCTACGAATCAACGG 2481
 751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLys 767
 2482 CTCGCGGCTGTGCTGACTGACAGCTTGTGGTAACTTGAAGAACTTACAA 2531

US-09-697-089-2 x AAH9581 ..
Align seq 1/1 to: AAH9581 from: 1 to: 2950

398 LeuGluGlyAlaPheSerHisLysPheAspPheGluLeuGlnAspValSe 414
|||||
47 CTGGAGGGTGTGTTCTCCCAAGATTGATTGCAACTGCAGAGATGTCTC 96
414 rSerValAsnGluAspValLeuLeuThrTrhGlyLeuLeuCysLysTrp 431
|||||
97 CACGCGAATGAGAGATGCTCTGCTGACAACTGGGCTCTCTGTAATATA 146
431 hraIaGlnArgPheLysProLysTrpLysPheHisLysSerPheGln 447
|||||
147 CACCTCAAGGTTCAAGCCAAAGTATTAATTTCTTTCACAGTCATCTCCAG 196
448 GluTrpThrAlaGlyArgGluSerSerLeuLeuThrSerHisGluTrp 464
|||||
197 GACTACACAGCAGCAGAGACTCAGCAGTTTATTGACGTCATGAGCC 246
464 OGluGluValThrLysGluValAsnGlyTrpLeuGlnLysMetValSerLys 481
|||||
247 AGAGGAGGTGACCAAGGGAGATGTTACTTGCAAGAAATGGTTCCATTT 296
481 eArspIleThrSerThrTySerSerLeuLeuArgTrpThrCysGlySer 497
|||||
297 CGGACATTACATCCACTTATAGCAGCCTGCTCGGTACACCTGGGTCA 346
498 SerValGluAlaThrArgAlaValMetLysHisLeuAlaAlaValTrpG 514
|||||
347 TCTGTGGAAGCCACAGGCGCTTATGAAACACCTGCACAGCTGATCA 396
514 nHISGlyCysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArg 531
|||||
397 ACACGGCTGCTCTCGACTTTCATGCCAAGAGGCTCTCTGGAGAC 446
531 InGluSerLeuGlnSerValLysAsnThrTrhGluGlnGluLeuLys 547
|||||
447 AGGAATCTTGGCAAGGTGAAAAACACCAGCAGCAAGAAATTCGAAA 496
548 AlaIleAsnIleAsnSerPheValGluCysGlyIleHisLeuTrpGln 564
|||||
497 GCATTAACATCAATTCCTTTCAGAGTGGCATTCATTATATCAGA 546
564 uSerThrSerLysSerAlaLeuSerGlnGluPheGluAlaPhePheGln 581
|||||
547 GAGTACATCCAAATTCAGCCCTGAGCCAGAAATTTGAAGCTTTCTTCAAG 596
581 LysSerLeuLeuTrpIleAsnSerGlyAsnIleProAspTrpLeuPheAsp 597
|||||
597 GTAAGAGCTATATATCAACTCAGGGAACATCCCGATTACTTATTGGAC 646
598 PhePheGluHisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLe 614
|||||
647 TTCCTTGAACATTTGGCCCAATTTGCAAGTGTCTGCACTTCAATTAAC 696
614 uAspPheTrpGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspTrp 631
|||||
697 GGGCTTTTATAGGGGAGCTATGGCTTCATGGAAGAGCTGCGAAGACACA 746
631 hrcIglyIleHisMetGluGluAlaProGluTrpTrpIleProSerArg 647
|||||
747 CAGGTGAATCCACATGGAAGAGGCCAGAAACCTACATTTCCAGAGAG 796
648 AlaValSerLeuPhePheAsnTrpLysGlnGluPheArgTrhLeuGluVal 664
|||||
797 GGTGTATCTTGTCTTCACTGGAAGCAGAAATTCAGGACTGTGGAGGT 846
664 LTrhLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTrpLeuG 681
|||||
847 CACACTCCGGGATTTTCAGCAAGTTGAATTAAGCAAGATATCAGATCTGC 896
681 LysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArgCys 697

897 GGAATAATATTCAGCTGTGCCAACAGCCTCAGGCTGCAATTAAGAGATGT 946
698 AlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsnI 714
|||||
947 GCTGTGTGTGGCTGGAACCTTCACGTTTGGTCTCAGCACCTGTAAAGCAT 996
714 eTrpSerLeuMetValGluAlaSerProLeuThrIleGluAspGluArg 731
|||||
997 TTATTTCTCTATGGTGGAAACCACTCCCTCACCATTGAAAGATGAGAGCC 1046
731 IsIleThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGln 747
|||||
1047 ACATCAGATCTGTAAACAAACCTGAAACCTTGAGTATTTCATGACCTCAG 1096
748 AsnGlnArgLeuProGlyLeuThrAspSerLeuGluLysLeuLysAs 764
|||||
1097 AATCAACGGCTGCCGGGTGTGACTGACAGCAGCTTGGGTAACTTGAAGAA 1146
764 nLeuThrLysLeuIleMetAspAsnIleLysMetAsnGluGluAspAla 781
|||||
1147 CCTTACAAAGCTCATATATGATTAACATTAAGATGATGAAGAGATGCTTA 1196
781 LeuLeuAlaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHis 797
|||||
1197 TAAAGTACAGTGAAGGCGCTGAAAACCTGAAAGAGATGTGTTATTTCAT 1246
798 LeuThrHisLeuSerAspIleGlyGluGlyMetAspTrpIleValLysSe 814
|||||
1247 TTGACCCACTTGTCTGTGACATTGAGAGGGAATGATTACATCTCAAGTC 1296
814 rLeuSerSerGluProCysAspLeuGluGluIleGlnLeuValSerCysC 831
|||||
1297 TCTGTCAAGTGAACCCCTGTACCTTGAAAGAAATTCATTTGCTCCTGCT 1346
831 yLeuSerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeu 847
|||||
1347 GCTTGTCTGCAAAATGCAAGTGAATAATCCTACCTCAGAAATCTTCAACAATTTG 1396
848 ValLysLeuSerIleLeuAspLeuSerGluAsnTrpLeuGluLysAspG 864
|||||
1397 GTCAAACTGAGCATTTCTGTATTATTCAGAAATTTACCTGGAAAAAGATGG 1446
864 yAsnGluAlaLeuHisGluLeuLeuIleAspArgMetAsnValLeuGluGln 881
|||||
1447 AAATGAAGCTTTCATGTAAGTCACTGATCCAGAGATGAAGCTGTGAGACAGC 1496
881 euthrAlaLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSer 897
|||||
1497 TCACCGCAGTGAATGCTGCCCTGGGGCTGTGACGTGCAAGCAGACCTGAGC 1546
898 SerLeuLeuLysHisLeuGluGluValProGlnLeuValLysLeuGlyLe 914
|||||
1547 AGCCTGTGAACATTTTGGAGAGGATCCCAACCTGCTCAAGCTTGGGCTT 1596
914 uLysAsnTrpArgLeuThrAspTrhGluIleArgIleLeuGlyAlaPheP 931
|||||
1597 GAAAACGTGAGACTCAGAGATACAGAGATTAGCAATTTTGGTGCAATTTT 1646
931 hecIlyAsnProLeuLysAsnPheGlnIleLeuAsnLeuAlaGlyAsn 947
|||||
1647 TTGGAAGAAGACCTTGAAAAACTTCACAGATTGAATTTGGCGGGAAT 1696
948 ArgValSerSerAspGlyTrpLeuAlaPheMetGlyAlaPheGluAsnLe 964
|||||
1697 CGTGTAGCAGTATGATGATGCTTGGCTTCATGGGTATTTAGAGATCT 1746
964 uLysGlnLeuValPhePheAspPheSerThrLysGluPheLeuProAsp 981
|||||
1747 TAAGCATTTAGTGTGTTTTCATCTTGTAGTAAAGAAATTTCTACCTGATC 1796
981 roaIaLeuValArgLysLeuSerGlnValLeuSerLysLeuThrPheLeu 997
|||||

```

1797 CAGCATTGCTCAGAAACTTAGCCAGTGTATCCAGTTAACTTTCTG 1846
998 GlnGluAlaArgLeuValGlyTrpGlnPheAspAspPLeuSerVa 1014
|||||
1847 CAGAGAGCTAGGCTTGTGGGTGCAATTTGATGATGATGATCAGTCT 1896
1014 lIleThrGlyAlaPheLysLeuValThrAla 1024
|||||
1897 TATTACAGTGTCTTTAAACTAGTAACTGCT 1927

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH34171
seq_documentation_block:
ID AAH34171 standard; cDNA; 2735 BP.
XX
XX AAH34171;
AC
XX 03-SEP-2001 (first entry)
DT
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1253.
DE
XX
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200122920-A2.
PN
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000MO-US26524.
PF
XX
XX 29-SEP-1999; 9905-0157137.
PR
XX 03-NOV-1999; 9905-0163280.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX
XX WPI; 2001-235357/24.
DR
XX
XX P-PSDB; AAG74766.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3017; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other;
XX

```

```

alignment_scores:
Quality: 180.00      Length: 180
Ratio: 1.000        Gaps: 0

```

```

Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AAH34171/rev
Align seg 1/1 to reverse of: AAH34171 from: 1 to: 2735
845 HisAsnLeuValLysLeuSerIleLeuAspLeuSerGluAsnTyrLeuG1 861
|||||
2727 CACAATTTGCTCAAACTGATGAGCATTTCTTGATTTATCAGAAATTAACCTGGA 2678
861 uLysAspGlyAsnGluAlaLeuHisGluLeuIleAspArgMetAsnVal1 878
|||||
2677 AAAAGATGGAATGAAAGCTCTTCATGAACTGATCAGACAGAGAAAGCTGC 2628
878 euGluGlnLeuThrAlaLeuMetLeuProTrrGlyCysAspValGlnGly 894
|||||
2627 TAGAACAGCTCACCCGACTGATGCTGCCCTGGGGCTGTGACCTGCAAGGC 2578
895 SerLeuSerSerLeuLeuLysHisLeuGluValProGlnLeuVal1y 911
|||||
2577 ACCCTGAGCAGCTTGTGTAACATTTGGAGAGGTCCCAACTGCTCA 2528
911 sLeuGlyLeuLysAsnTrrArgLeuThrAspThrGluIleArgIleLeuG 928
|||||
2527 GCTTGGGTGAAAAACTGAGACTCAGATCAGATCAGATTAATTTTAA 2478
928 lYAlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeu 944
|||||
2477 GTGCATTTTGTGAAAGAACCTCTGAAAAAACCCTCCAGCAGTGAATTTG 2428
945 AlaGlyAsnArgValSerSerAspGlyTrrLeuAlaPheMetGlyValPh 961
|||||
2427 GCGGAAATCGTGTGAGCAGTATGATGATGCTTGCCTTCATGGGTAT 2378
961 eGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPheL 978
|||||
2377 TGAGAACTTTAAGCAATTAAGTGTGTTTTCACCTTAGTACTAAAGAAATTC 2328
978 euProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeu 994
|||||
2327 TACCTGATCCACATTAAGTCAAGAAACTTAGCCAGTGTATCCAAAGTTA 2278
995 ThrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAspAs 1011
|||||
2277 ACTTTTCTGCAGAAAGCTAGCTGTGGTGGCAATTTGATGATGATGATGA 2228
1011 PLeuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
|||||
2227 TCTCAGTGTATTACAGGTGCTTTTAAACTAGTAACTGCT 2188

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI14389
seq_documentation_block:
ID AAI14389 standard; DNA; 421 BP.
XX
XX AAI14389;
AC
XX 12-OCT-2001 (first entry)
DT
XX
XX Probe #4322 for gene expression analysis in human cervical cell sample.
DE
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KM cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001MO-US00670.
XX

```

PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 25; SEQ ID NO 4322; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcf_sequences.
XX
SQ Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

alignment_scores:
Quality: 80.00 Length: 80
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAI14389 ..

Align seg 1/1 to: AAI14389 from: 1 to: 421

928 GLYAlaPheheGlyLysAsnProLeuLysAsnPhelngInLeuAsnle 944
|||||
181 GGTCATTTTGGAAAGAACCTCTGAAACTTCCAGCAGTTGAAATT 230
944 uAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyValP 961
|||||
231 GCGGGGAAATCGTGTGAGCAGTGTGATGCTTGCCTTCATGGGTAT 280
961 heGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPhe 977
|||||
281 TTGAGATCTTTAAGCAATTAAGTCTTTTTCACCTTACTTAAGAAATTT 330
978 LeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysle 994
|||||
331 CTACCTGATCCAGCATTAAGTACAGAAACTTACGCCAAGTGTATCCAGTT 380
994 utrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPhe 1007
|||||
381 AACCTTTCTGCAGAGCTAGCTGTGGTGGCAATTT 420

seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA2001.DAT:AAI135764

seq_documentation_block:
ID AAI135764 standard; DNA; 421 BP.
XX
AC AAI135764;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #4450 used to measure gene expression in human placenta sample.

XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
PS Claim 25; SEQ ID NO 4450; 654pp; English.
XX
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

alignment_scores:
Quality: 80.00 Length: 80
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAI135764 ..

Align seg 1/1 to: AAI135764 from: 1 to: 421

928 GLYAlaPheheGlyLysAsnProLeuLysAsnPhelngInLeuAsnle 944
|||||
181 GGTCATTTTGGAAAGAACCTCTGAAACTTCCAGCAGTTGAAATT 230
944 uAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyValP 961
|||||
231 GCGGGGAAATCGTGTGAGCAGTGTGATGCTTGCCTTCATGGGTAT 280
961 heGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPhe 977
|||||
281 TTGAGATCTTTAAGCAATTAAGTCTTTTTCACCTTACTTAAGAAATTT 330
978 LeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysle 994
|||||
331 CTACCTGATCCAGCATTAAGTACAGAAACTTACGCCAAGTGTATCCAGTT 380
994 utrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPhe 1007
|||||
381 AACCTTTCTGCAGAGCTAGCTGTGGTGGCAATTT 420

seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA2001.DAT:AAI04213

seq_documentation_block:

```

ID  AA104213 standard; DNA; 421 BP.
XX
AC  AA104213:
XX
DT  09-OCT-2001 (first entry)
XX
DE  Probe #4204 used to measure gene expression in human breast sample.
XX
KW  Probe: human; breast disease; breast cancer; development disorder; ss;
XX  inflammatory disease; proliferative breast disease; non-carcinoma tumour.
OS  Homo sapiens.
XX
XX  WO200157270-A2.
XX
PD  09-AUG-2001.
XX
PF  29-JAN-2001; 2001WO-US00661.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX  WPI; 2001-476286/51.
XX
PT  Novel single exon nucleic acid probe used to measuring gene expression
XX  in a human breast.
XX
PS  Claim 25; SEQ ID No 4204; 322pp; English.
XX
CC  The present invention relates to novel single exon nucleic acid probes.
CC  The present sequence is one such probe. The probes are useful for
CC  measuring human gene expression in a human breast sample, where the probe
CC  hybridises at high stringency to a nucleic acid expressed in the human
CC  breast. The probes are useful for predicting, diagnosing, grading,
CC  staging, monitoring and prognosing diseases of the human breast,
CC  particularly those diseases with polygenic aetiology. The diseases
CC  include: breast cancer, disorders of development, inflammatory diseases
CC  of the breast, fibrocystic changes, proliferative breast disease and
CC  non-carcinoma tumours.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

alignment_scores:
    Quality: 80.00      Length: 80
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA104213
Align seg 1/1 to: AA104213 from: 1 to: 421

```

```

961  heGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPhe 977
      |||||||
281  TTGAGATCTTAAAGCAATTAGTCTTTTGGACTTGTACTAAAGATT 330
      |||||||
978  LeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLe 994
      |||||||
331  CTACCTGATCCACGATTAGTCAGAAACCTTACCAAGTGTATCCCAAGTT 380
      |||||||
994  uThrPheLeuGlnGlnLAlaArgLeuValGlyTrpGlnPhe 1007
      |||||||
381  AACTTTCTGCAGAGAGCTAGGCTTGTGGTGCAATT 420

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AA123590
seq_documentation_block:
ID  AA123590 standard; DNA; 220 BP.
XX
AC  AA123590;
XX
DT  12-OCT-2001 (first entry)
XX
DE  Probe #13523 for gene expression analysis in human cervical cell sample.
XX
KW  Probe: human; microarray; gene expression; cervical epithelial cell;
XX  cervical cancer; ss.
XX
OS  Homo sapiens.
XX
XX  WO200157278-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00670.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX  WPI; 2001-488901/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
XX  analyzing gene expression in human cervical epithelial cells.
XX
PS  Claim 25; SEQ ID No 13523; 487pp; English.
XX
CC  The present invention relates to human single exon nucleic acid probes
CC  (SENPs). The present sequence is one such probe. The SENPs are derived
CC  from human HeLa cells. The SENPs can be used to produce a single exon
CC  microarray, which can be used for measuring human gene expression in a
CC  sample derived from human cervical epithelial cells. By measuring gene
CC  expression, the probes are therefore useful in grading and/or staging
CC  of diseases of the cervix, notably cervical cancer.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

alignment_scores:
    Quality: 73.00      Length: 73
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:

```

us-09-697-089-2 x AA123590 ..

Align seg 1/1 to: AA123590 from: 1 to: 220

```

944 LeuAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyVa 960
|||||
1 TTGGCGGGAATCGTGTGACAGTGAATGATGATGCTTCCATTCATGGGTGT 50
960 lPheGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGlu 977
|||||
51 ATTTGAGAAATCTTAAGCAATTAAGTGTGTTTGTGACTTGTAGTAAAGAAAT 100
977 heLeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLys 993
|||||
101 TTCTACTGATCCAGCATTTAGTCAGAAACTTACCCAGAGTGTATCCAG 150
994 LeuThrPheLeuGlnGlnLualArgLeuValGlyTrpGlnPheAspAsp 1010
|||||
151 TTTAACTTTCTGCAGAGAGCTAGGCTTGTGGTGGCAATTTGATGATGA 200
1010 PaSplLeuSerValIleThr 1016
|||||
201 TGATCTCAGTGTATTACA 219

```

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AA148904

seq_documentation_block:

ID AA148904 standard; DNA: 220 BP.

AC AA148904;

DT 17-OCT-2001 (first entry)

DE Probe #17590 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

KM genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 17590; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

alignment_scores:

Quality: 73.00

Length: 73

Ratio: 1.000

Gaps: 0

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

us-09-697-089-2 x AA148904 ..

Align seg 1/1 to: AA148904 from: 1 to: 220

```

944 LeuAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyVa 960
|||||
1 TTGGCGGGAATCGTGTGACAGTGAATGATGATGCTTCCATTCATGGGTGT 50
960 lPheGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGlu 977
|||||
51 ATTTGAGAAATCTTAAGCAATTAAGTGTGTTTGTGACTTGTAGTAAAGAAAT 100
977 heLeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLys 993
|||||
101 TTCTACTGATCCAGCATTTAGTCAGAAACTTACCCAGAGTGTATCCAG 150
994 LeuThrPheLeuGlnGlnLualArgLeuValGlyTrpGlnPheAspAsp 1010
|||||
151 TTTAACTTTCTGCAGAGAGCTAGGCTTGTGGTGGCAATTTGATGATGA 200
1010 PaSplLeuSerValIleThr 1016
|||||
201 TGATCTCAGTGTATTACA 219

```

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AA109206

seq_documentation_block:

ID AA109206 standard; DNA: 220 BP.

AC AA109206;

DT 09-OCT-2001 (first entry)

DE Probe #9197 used to measure gene expression in human breast sample.

KW Probe; human; breast disease; breast cancer; development disorder; ss;

KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

PE 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-476286/51.

PT Novel single exon nucleic acid probe used to measuring gene expression

PT in a human breast -

PS Claim 25; SEQ ID No 9197; 322pp; English.

CC The present invention relates to novel single exon nucleic acid probes.

CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast.
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_sequences.
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

alignment_scores:
Quality: 73.00 Length: 73
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA109206 ..

Align seg 1/1 to: AA109206 from: 1 to: 220

944 LeuAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyVa 960
1 TTGGGGGAAATCGTGAGCAGTCATGATGCTTGGCTTCATGGGCGT 50
960 lPheGlnuSneUySgInleuValPhePheAspPheSerThrIlySgIuP 977
51 ATTGAGATCTTAAAGCAATAGTCTTTTTCACCTTACTTAAAGAAAT 100
977 heLeuPcAspPcAlaLeuValArgLysLeuSerGlnValLeuSerLys 993
101 TTCTACCTGCATCCAGCATTCAGCAAAAACCTTAGCCAACTGTATCCAA 150
994 LeuThrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAspAs 1010
151 TTAACTTTTCTGCAGAAAGCTAGCTGTGGTGCAATTGATGATGATGA 200
1010 PASpleuSerValIleThr 1016
201 TGATCTCAGTGTATTACA 219

seq_name: /SIDIS2/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC54488
seq_documentation_block:
ID AAC54488 standard; DNA: 848 BP.
XX
AC AAC54488;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78048.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131448.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134378.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.


```

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

```

```

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

```

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-697-089-2 x AAC54488/rev ..

```

```

Align seg 1/1 to reverse of: AAC54488 from: 1 to: 848

```

```

239 LeuArgGlnArgValLeuPheLeu 247
|||||
601 CTCGGCAAGAGCTCTTCTTCTCTT 575

```

```

seq_name: /SIDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAC54274

```

```

seq_documentation_block:

```

```

ID AAC54274 standard; DNA; 850 BP.

```

```

AC AAC54274;

```

```

XX 18-OCT-2000 (first entry)

```

```

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77302.

```

```

KW Hybridisation assay; genetic mapping; gene expression control;

```

```

KW protein identification; signal transduction pathway;

```

```

KW metabolic pathway; promoter; termination sequence; ss.

```

```

OS Arabidopsis thaliana.

```

```

XX EP1033405-A2.

```

```

XX 06-SEP-2000.

```

```

PF 25-FEB-2000; 2000EP-0301439.

```

```

XX 25-FEB-1999; 99US-0121825.

```

```

PR 05-MAR-1999; 99US-0123180.

```

```

PR 09-MAR-1999; 99US-0123548.

```

```

PR 23-MAR-1999; 99US-0125788.

```

```

PR 25-MAR-1999; 99US-0126264.

```

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145210.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.
```

```
alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000
```

alignment_block:

US-09-697-089-2 x AAC54274/rev ...

Align seg 1/1 to reverse of: AAC54274 from: 1 to: 850

```
239 LeuArgGlnArgValLeuPheLeu 247
|||||
603 CTCGCGCAAGAGCTTCTTCTCTT 577
```

seq_name: /STD2/gcgcdata/geneseq/geneseqn/NA2001.DAT:AAH94390

seq_documentation_block:

ID AAH94390 standard; cDNA; 896 BP.

AC AAH94390;

DT 05-OCT-2001 (first entry)

DE Human foetal cDNA, SEQ ID NO: 919.

XX Human: foetal protein; cytostatic; immunosuppressive; immunostimulant;

KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;

KW gene therapy; antisense therapy; cancer; immune disorder;

KW growth disorder; osteoporosis; thrombolytic disorder;

KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.

XX Homo sapiens.

OS WO200155339-A2.

PN 02-AUG-2001.

XX 25-JAN-2001; 2001MO-US02723.

XX 25-JAN-2000; 2000US-0491404.

PR 15-SEP-2000; 2000US-0663870.

PR 06-NOV-2000; 2000US-0707351.

PA (HYSE-) HYSEQ INC.

XX Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;

PI Liu C, Asundi V, Zhou P, Werhman T;

XX WPI; 2001-465571/50.

DR P-PSDB; AAM06715.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases

PT associated with dysfunction of the protein e.g. cancers, immune

PT disorders, growth disorders, thrombolytic disorders, nervous system

PT disorders and inflammation -

PS Claim 1; Page 538; 715pp: English.

CC The invention relates to novel foetal polypeptides encoded by

CC polynucleotides comprising one of 477 sequences fully defined in the

CC specification. The foetal polynucleotides and polypeptides are

CC useful in the treatment and diagnosis of diseases such as cancers,

CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic

CC disorders, nervous system disorders and inflammation. The present

CC sequence is a full length cDNA which was assembled using expressed

CC sequence tags (ESTs) found to be expressed in human foetal tissue

CC cDNA libraries as seeds.

SQ Sequence 896 BP; 183 A; 250 C; 272 G; 187 T; 4 other;

```
alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000
```

alignment_block:

US-09-697-089-2 x AAH94390/rev ..

Align seg 1/1 to reverse of: AAH94390 from: 1 to: 896

```
450 ThrAlaGlyArgArgLeuSerLeu 458
|||||
722 ACTGCTGCGAAGAGCTCTCTCCCTA 696
```

seq_name: /STD2/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAZ36643

seq_documentation_block:

ID AAZ36643 standard; cDNA; 1382 BP.

AC AAZ36643;

DT 22-FEB-2000 (first entry)

DE Human tumour rejection antigen RUR-1 antisense cDNA sequence.

XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;

KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;

KW leukaemia; ss.

XX Homo sapiens.

OS Key

FT CDS

FT primer_bind

FT primer_bind

PN WO958546-A1.

PD 18-NOV-1999.

```

PF 13-MAY-1999; 99WO-US10424.
XX
PR 13-MAY-1998; 98US-0085318.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-053076/04.
DR P-PSDB; AAY53809.
XX
PT New isolated tumour rejection antigen RUR-1 nucleic acids, used for,
PT e.g. treatment of cancers -
XX
PS Claim 4; Fig 5; 75pp; English.
XX
CC The present sequence represents the antisense cDNA sequence of human
CC tumour rejection antigen RUR-1. The present sequence is the antisense
CC strand of a ubiquitously expressed gene. The antisense strand codes for
CC a polypeptide which is preferentially expressed in tumour samples and
CC tumour-derived cell lines. The polypeptide is unrelated to any TRAP
CC protein. The sequence was isolated from a renal cell carcinoma line
CC l89211-RCC. The RUR-1 nucleic acids and polypeptides can be used for
CC diagnosis, prognosis or treatment of a disorder characterized by the
CC expression of a RUR-1 antisense cDNA molecule or an expression product,
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,
CC melanoma, sarcoma or leukaemia.
XX
SO Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 other;

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA236643 ..

Align seg 1/1 to: AA236643 from: 1 to: 1382

451 AlaGlyArgArgLeuSerLeuLeu 459
      |||||||
856 GCGGCGCGCGCTCAGCTGCTT 882

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AA236548
seq_documentation_block:
ID AAC39548 standard; DNA; 1854 BP.
XX
AC AAC39548;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25036.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.

```

```

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

```

PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 03-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-015370.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.

PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 13-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161922.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

alignment_scores: Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAC39548/rev ..

Align seg 1/1 to reverse of: AAC39548 from: 1 to: 1854

762 LeuLysAsnLeuThrIlystLeuIleMet 770

|||||
1798 TTTAAAAATCTTACAAACATATTATG 1772

seq_name: /SIDS7/gcgcdata/geneseq/geneseqn/NA2001.DAT:AA158192

seq_documentation_block:

ID AA158192 standard; CDNA; 1956 BP.

AC AA158192;

DT 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 395.

DE Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX leukemia; ss.

OS Homo sapiens.

XX WO200153312-A1.

PD 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.

```

PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0634350.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSED INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM39036.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 395; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, Leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1956 BP; 540 A; 435 C; 541 G; 435 T; 5 other:

Alignment_scores:
      Quality: 9.00      Length: 9
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA158192/rev ..

Align seg 1/1 to reverse of: AA158192 from: 1 to: 1956

451 AAGIYARGARGLeuSerSerLeuLeu 459
|||||
293 GCTGGCGCGCGCTCAGCTCGCTGCTT 267

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AA236644
seq_documentation_block:
ID AA236644 standard; cDNA; 2167 BP.
XX
AC AA236644;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human tumour rejection antigen RUR-1 sense cDNA sequence.
XX
KM Antisense: human; tumour rejection antigen: RUR-1; tumour; cancer;
KM renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;
KM leukaemia; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 303..1733

```

```

FT primer_bind
FT 484..501
FT /*tag= a
FT /*tag= b
FT /*note= "binding site for primer VDE87 (see AA236645)"
FT primer_bind
FT complement (917..935)
FT /*tag= C
FT /*note= "binding site for primer VDE93 (see AA236646)"
XX
PN WO958546-A1.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10424.
XX
PR 13-MAY-1998; 98US-0085318.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Van Den Eynde B, Boon-Falleur T;
XX
DR WPI: 2000-053076/04.
DR P-PSDB; AAY43811.
XX
PT New isolated tumour rejection antigen RUR-1 nucleic acids, used for,
PT e.g. treatment of cancers -
XX
PS Claim 5; Fig 6; 75pp; English.
XX
CC The present sequence represents the sense cDNA sequence of a human
CC ubiquitously expressed gene. The antisense strand of the present
CC sequence encodes tumour rejection antigen RUR-1. The antisense strand
CC codes for a polypeptide which is preferentially expressed in tumour
CC samples and tumour-derived cells lines. The polypeptide is unrelated
CC to any TRAP protein. The antisense sequence was isolated from a renal
CC cell carcinoma line 189211-RCC. The RUR-1 nucleic acids and polypeptides
CC can be used for diagnosis, prognosis or treatment of a disorder
CC characterized by the expression of a RUR-1 antisense cDNA molecule or
CC an expression product, such as cancers, e.g. renal cell carcinoma,
CC colorectal carcinoma, melanoma, sarcoma or leukaemia.
XX
SQ Sequence 2167 BP; 675 A; 440 C; 576 G; 476 T; 0 other:

Alignment_scores:
      Quality: 9.00      Length: 9
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA236644/rev ..

Align seg 1/1 to reverse of: AA236644 from: 1 to: 2167

451 AAGIYARGARGLeuSerSerLeuLeu 459
|||||
295 GCTGGCGCGCGCTCAGCTCGCTGCTT 269

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:AA14274
seq_documentation_block:
ID AA14274 standard; DNA; 3166 BP.
XX
AC AA14274;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1661 gene.
XX
KM GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX

```

```

FH Key Location/Qualifiers
FT CDS 51..316
/*tag- a
XX
XX W09843478-A1.
XX
XX 08-OCT-1998.
XX
XX 01-APR-1998: 98MO-US06371.
XX
XX 29-JUL-1997: 97US-0902615.
XX 01-APR-1997: 97US-0833457.
XX 24-JUN-1997: 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
XX
XX WPI: 1998-542293/46.
XX
XX P-PSDB; AAW98555.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
XX for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
XX
XX Claim 1; Page 1082-1087; 2054pp; English.
XX
XX This sequence represents a polynucleotide of the invention. It was
XX isolated from Helicobacter pylori and encodes a H.pylori GHP0 protein.
XX The polypeptides can be used for preventing or treating Helicobacter
XX infections, and gastroduodenal diseases associated with these
XX infections, including acute, chronic, and atrophic gastritis, and peptic
XX ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
XX for the production of antibodies. The products can also be used for
XX detection and diagnosis.
XX
XX Sequence 3166 BP; 1146 A; 448 C; 622 G; 950 T; 0 other;
XX
XX alignment_scores:
XX Quality: 9.00 Length: 9
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-697-089-2 x AAX14274 ..
XX
XX Align seg 1/1 to: AAX14274 from: 1 to: 3166
XX
XX 735 ValThrAsnLeuLysThrLeuSerIle 743
XX ||||||||||||||||||||||||||||
XX 2685 GTTACTAAGCTTAAGACTTTAAGTATT 2711
XX
XX seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT: AAX12965
XX
XX seq_documentation_block:
XX ID AAX12965 standard; DNA: 7515 BP.
XX
XX AAX12965;
XX
XX 19-MAR-1999 (first entry)
XX
XX Enterococcus faecalis genome contig SEQ ID NO:28.
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
XX
XX W09850555-A2.
XX
XX 12-NOV-1998.
XX
```

```

XX
XX 04-MAY-1998: 98MO-US08985.
XX
XX 14-NOV-1997: 97US-0066009.
XX 06-MAY-1997: 97US-0044031.
XX 16-MAY-1997: 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon PJ, Kunsch CA;
XX
XX WPI: 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
XX - used to develop products for the detection of Enterococcus and for
XX use in vaccines for prevention or attenuation of Enterococcus
XX infection.
XX
XX Claim 1; Page 380-384; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
XX 962 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX AAX12938 to AAX13919 represent these nucleotide sequences which are
XX primary nucleotide sequences, also known as contigs. The computer-based
XX system can identify fragments of the Enterococcus faecalis genome with
XX commercial importance. The products can be used to detect the presence
XX of Enterococcus faecalis in samples. They can also be used for
XX diagnosing Enterococcal infection in an animal and monitoring
XX progression of disease, and for identifying agents which can be used to
XX modulate the growth or pathogenicity of Enterococcus faecalis, or
XX another related organism, in vivo or in vitro. In particular the
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX can be used in vaccines to prevent or attenuate an Enterococcal
XX infection.
XX
XX Sequence 7515 BP; 2118 A; 1566 C; 1281 G; 2543 T; 7 other;
XX
XX alignment_scores:
XX Quality: 9.00 Length: 9
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-697-089-2 x AAX12965/rev ..
XX
XX Align seg 1/1 to reverse of: AAX12965 from: 1 to: 7515
XX
XX 236 LeuLeuLysLeuArgGlnArgValIleu 244
XX ||||||||||||||||||||||||||||
XX 5470 TTGTTGAAATTCGCGACACGCGTCTCG 5444
XX
XX seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT: AAS01960
XX
XX seq_documentation_block:
XX ID AAS01960 standard; DNA: 249 BP.
XX
XX AAS01960;
XX
XX 04-JUL-2001 (first entry)
XX
XX Cytochrome P-450 (CYP)3A4 gene fragment containing exon 3.
XX
XX CYP3A4; CYP3A7; human; exon/intron boundary; cytochrome P-450; cancer;
XX abnormal drug response; environmental carcinogen; genotype; polymorphism;
XX drug candidate; protein malfunction; inhibitor; hypersensitivity; ds;
XX
XX Homo sapiens.
XX
XX key Location/Qualifiers
XX FH 1..115
XX FT intron
XX /*tag- a
XX
```

```

FT FT /cons_splice= (5'site:NO,3'site:YES)
FT CDS 116..166
FT /tag= b
FT /partial
FT /product= "Human CYP3A4 fragment"
FT /note= "No start and stop codon"
FT exon 116..168
FT /tag= c
FT Intron 169..249
FT /tag= d
FT /cons_splice= (5'site:NO,3'site:NO)
FT
FT
FT WO200120025-A2.
FT
FT
FT 22-MAR-2001.
FT
FT
FT 01-SEP-2000; 2000WO-EP08570.
FT
FT
FT 10-SEP-1999; 99EP-0118120.
FT
FT
FT (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
FT
FT
FT PA
FT XX
FT PI Wojnowski L, Eiselt R;
FT
FT DR WPI: 2001-244818/25.
FT
FT DX P-PSDB; AA000990.
FT
FT
FT Novel variant of CYP3A4 and CYP3A7 genes, associated with insufficient
FT metabolism and/or sensitivity to drugs, useful for diagnosing and
FT treating diseases with drugs that are modulators of their gene product
FT
FT PT
FT
FT XX
FT PS
FT Claim 1; Fig 6; 106pp; English.
FT
FT
FT The sequence represents a fragment of the cytochrome P-450 (CYP)3A4 gene
FT containing exon 3. Polymorphic polynucleotides of the CYP3A4 or CYP3A7
FT genes are associated with abnormal drug response or individual
FT
FT CC predisposition to several common cancers caused by environmental
FT carcinogens. The primer sequences can be used in the production of
FT variant CYP3A4 and CYP3A7 proteins in order to study the malfunction of
FT the proteins, and in diagnostic tests designed for the specific detection
FT CC and genotyping of CYP3A4 and CYP3A7 alleles in humans. The invention
FT provides methods for identifying and obtaining drug candidates and
FT CC inhibitors of the genes for therapy of disorders related to acquired drug
FT hypo- or hypersensitivity.
FT
FT
FT XX
FT Sequence 249 BP; 69 A; 47 C; 51 G; 81 T; 1 other;
FT
FT
FT alignment_scores:
FT
FT Quality: 8.00 Length: 8
FT Ratio: 1.000 Gaps: 0
FT Percent Similarity: 100.000 Percent Identity: 100.000
FT
FT
FT alignment_block:
FT US-09-697-089-2 x AAS01960/rev ..
FT
FT
FT Align seg 1/1 to reverse of: AAS01960 from: 1 to: 249
FT
FT 361 ThrLeuPheHisThrPheTyrAsp 368
FT |||||||
FT 164 ACACTTTTCCATCACTTTTATGAC 141
FT
FT seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT.AAH36261
FT
FT seq_documentation_block:
FT ID_AAH36261 standard; CDNA; 287 BP.
FT
FT
FT AAH36261;
FT
FT
FT 03-SEP-2001 (first entry)
FT
FT
FT Human colon cancer antigen encoding CDNA SEQ ID NO:3343.
FT
FT

```

```

XX XX Human: colon cancer; colon cancer antigen; diagnosis; detection:
KW KM colorectal carcinoma; ss.
XX OS Homo sapiens.
XX PN MO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WC-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX P1 WPI: 2001-235357/24.
XX DR P-PSDB; AAG76856.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PS useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 1; Page 5149-5150; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patient's own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX CC and AAB77789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX CC
XX CC Sequence 287 BP; 85 A; 47 C; 81 G; 61 T; 13 other;
XX
XX Alignment_scores:
XX Quality: 8.00 Length: 8
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX Alignment_block:
XX US-09-697-089-2 x AAH36261/rev ..
XX
XX Align seg 1/1 to reverse of: AAH36261 from: 1 to: 287
XX
XX 689 SerLeuArgLeuGlnIleIleTysArg 696
XX |||||||||||||||||||||||||
XX 70 TCCCTGAGGCTACAGATTAAAGAGC 47
XX
XX seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF08944
XX
XX seq_documentation_block:
XX ID AAF08944 standard; CDNA; 325 BP.
XX AC AAF08944;
XX AT
XX TT 13-MAR-2001 (first entry)
XX

```



```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA126514
seq_documentation_block:
ID   AA126514 standard; DNA; 347 BP.
XX
AC   AA126514;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Probe #16447 for gene expression analysis in human cervical cell sample.
XX
KW   Probe: human; microarray; gene expression; cervical epithelial cell;
KM   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
XX   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 16447; 487bp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SNP). The present sequence is one such probe. The SNPs are derived
CC   from human HeLa cells. The SNPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SO   Sequence 347 BP; 55 A; 111 C; 91 G; 90 T; 0 other;

alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
US-09-697-089-2 x AA126514  ..

Align seg 1/1 to: AA126514 from: 1 to: 347

316 GluteralaglyleuLeu 323
|||||
296 GAACCTGCAGAGCTCTCTT 319

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA155266
seq_documentation_block:
```

```
ID   AA155266 standard; DNA; 347 BP.
XX
AC   AA155266;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Probe #23952 used to measure gene expression in human placenta sample.
XX
KW   Probe: microarray; human; placenta; antenatal diagnosis;
KM   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-48897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
XX   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 23952; 654bp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SNP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SO   Sequence 347 BP; 55 A; 111 C; 91 G; 90 T; 0 other;

alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
US-09-697-089-2 x AA155266  ..

Align seg 1/1 to: AA155266 from: 1 to: 347

316 GluteralaglyleuLeu 323
|||||
296 GAACCTGCAGAGCTCTCTT 319

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH12387
seq_documentation_block:
ID   AAH12387 standard; cDNA; 375 BP.
XX
AC   AAH12387;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9222.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
```

```

XX OS Homo sapiens.
XX XX
XX PN EP1074617-A2.
XX PD
XX PD 07-FEB-2001.
XX XX
XX PF 28-JUL-2000; 2000EP-0116126.
XX XX
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX XX
XX PA (HELI-) HELIX RES INST.
XX XX
XX PI Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX XX
XX DR WPI; 2001-318749/34.
XX XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS
XX PS Claim 3: SEQ ID 9222; 2537pp + CD ROW; English.
XX XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX XX
XX SQ Sequence 375 BP; 123 A; 74 C; 45 G; 127 T; 6 other;
XX XX
XX Alignment_scores:
XX Quality: 8.00 Length: 8
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX XX
XX Alignment_block:
XX US-09-697-089-2 x AAH12387 ..
XX XX
XX Align seg 1/1 to: AAH12387 from: 1 to: 375
XX XX
XX 895 SerLeuSerSerLeuLeuLeuHis 902
XX ||||||||||||||||||||||||
XX 79 TCTTTATCAAGCCTTCCTTAAACAC 102
XX XX
XX seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF65760
XX seq_documentation_block:
XX ID AAF65760 standard; cDNA: 383 BP.

```

```

XX AAF65760;
XX
XX 09-APR-2001 (first entry)
XX
XX Novel human polynucleotide, SEQ ID NO: 1516.
XX
XX Human; cytosstatic; gene therapy; colon cancer; prostate cancer;
XX breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
XX
XX WO2001025568-A2.
XX
XX 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18374.
XX
XX 02-JUL-1999; 99US-0142310.
XX
XX 02-JUL-1999; 99US-0142311.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
XX Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
XX Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
XX Kita D, Garcia V, Jones LM, Strache-Crain B;
XX
XX WPI: 2001-091805/10.
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
XX mammalian cell and detecting cancer, particularly of the colon or
XX prostate, comprises 3351 human polynucleotide sequences -
XX
XX Claim 9; Page 761; 1046pp; English.
XX
XX The present sequence is one of 3351 sequences in a library of human
XX polynucleotides. The library is used to detect differentially expressed
XX genes correlated with a cancerous state of a mammalian cell and can
XX detect colon, prostate, breast and lung cancer. The library can be used
XX to produce probes for detection of mRNA and to produce additional copies
XX of the polynucleotides. The probes can be used for chromosome mapping or
XX the polynucleotide and for detection of transcription levels. Ribozymes
XX or antisense oligonucleotides can be generated. The polynucleotides and
XX their gene products are used as genetic or biochemical markers (e.g. in
XX blood or tissues) that will detect the earliest changes along the
XX carcinogenesis pathway and/or monitor the efficacy of therapies and
XX preventive interventions. The polynucleotides, polypeptides and
XX antibodies against them can be used in pharmaceutical compositions to
XX treat the cancers and proliferative disorders such as neoplasia,
XX dysplasia and hyperplasia.
XX
XX Sequence 383 BP; 103 A; 78 C; 105 G; 97 T; 0 other;
XX
XX
XX
XX
XX alignment_scores:
XX      Quality:      8.00      Length:      8
XX      Ratio:      1.000      Gaps:      0
XX Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-697-089-2 x AAF65760/rev ..
XX
XX Align seg 1/1 to reverse of: AAF65760 from: 1 to: 383
XX
XX 689 SerLeuArgLeuGlnIleLeuArg 696
XX |||||||
XX 94 TCCCTGAGCGCTACAGATTAAAGAG 71
XX
XX seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AA113139
XX
XX seq_documentation_block:

```

```
ID      AA113139 standard; DNA: 392 BP.
XX
XX      AA113139;
AC
XX      12-OCT-2001 (first entry)
DT
XX      Probe #3072 for gene expression analysis in human cervical cell sample.
DE
XX      Probe; human; microarray; gene expression; cervical epithelial cell;
KW      cervical cancer; ss.
XX
XX      Homo sapiens.
OS
XX      WO200157278-A2.
PN
XX      09-AUG-2001.
PD
XX      30-JAN-2001; 2001WO-US00670.
PE
XX      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX      WPI: 2001-488901/53.
DR
XX      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human cervical epithelial cells -
XX
XX      Claim 25; SEQ ID No 3072; 487pp; English.
PS
XX      The present invention relates to human single exon nucleic acid probes
CC      (SENPs). The present sequence is one such probe. The SENPs are derived
CC      from human HeLa cells. The SENPs can be used to produce a single exon
CC      microarray, which can be used for measuring human gene expression in a
CC      sample derived from human cervical epithelial cells. By measuring gene
CC      expression, the probes are therefore useful in grading and/or staging
CC      of diseases of the cervix, notably cervical cancer.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 392 BP; 107 A; 87 C; 118 G; 80 T; 0 other;
SQ

alignment_scores:
      Quality:      8.00      Length:      8
      Ratio:      1.000      Gaps:      0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA113139/rev ..

Align seg 1/1 to reverse of: AA113139 from: 1 to: 392

      316 GluLeuAlaGluGlyLeuLeu 323
      ||||||||||||||||||||||||
      174 GAACCTTGCAAGAGCTGCTCTT 151

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA134491

seq_documentation_block:
ID      AA134491 standard; DNA: 392 BP.
XX
XX      AA134491;
AC
XX
```

```
DT      17-OCT-2001 (first entry)
XX
XX      Probe #3177 used to measure gene expression in human placenta sample.
DE
XX      Probe; microarray; human; placenta; antenatal diagnosis;
KW      genetic disorder; ss.
XX
XX      Homo sapiens.
OS
XX      WO200157272-A2.
PN
XX      09-AUG-2001.
PD
XX      30-JAN-2001; 2001WO-US00663.
PE
XX      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX      WPI: 2001-48897/53.
DR
XX      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human placenta -
XX
XX      Claim 25; SEQ ID No 3177; 654pp; English.
PS
XX      The present invention relates to single exon nucleic acid probes (SENPs).
CC      The present sequence is one such probe. The probes are useful for
CC      producing a microarray for predicting, measuring and displaying gene
CC      expression in samples derived from human placenta. The probes are useful
CC      for antenatal diagnosis of human genetic disorders.
XX
XX      Sequence 392 BP; 107 A; 87 C; 118 G; 80 T; 0 other;
SQ

alignment_scores:
      Quality:      8.00      Length:      8
      Ratio:      1.000      Gaps:      0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA134491/rev ..

Align seg 1/1 to reverse of: AA134491 from: 1 to: 392

      316 GluLeuAlaGluGlyLeuLeu 323
      ||||||||||||||||||||||||
      174 GAACCTTGCAAGAGCTGCTCTT 151

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA103044

seq_documentation_block:
ID      AA103044 standard; DNA: 392 BP.
XX
XX      AA103044;
AC
XX      09-OCT-2001 (first entry)
DT
XX      Probe #3035 used to measure gene expression in human breast sample.
DE
XX      Probe; human; breast disease; breast cancer; development disorder; ss;
KW      inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX      Homo sapiens.
OS
XX
```

```

PN WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 25; SEQ ID NO 3035; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 392 BP; 107 A; 87 C; 118 G; 80 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA103044/rev ..

Align seg 1/1 to reverse of: AA103044 from: 1 to: 392

316 GluLeuAlaGluGlyLeuLeuLeu 323
|||||
174 GAACCTGCAGCAAGCTGCTCTCTT 151

seq_name: /STD2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH12192

seq_documentation_block:
ID AAH12192 standard; cDNA: 394 BP.
XX
XX AAH12192;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (3'-primer) SEQ ID NO:9027.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX BP1074617-A2.
XX

```

```

PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 3; SEQ ID 9027; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 394 BP; 106 A; 82 C; 79 G; 118 T; 9 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAH12192/rev ..

Align seg 1/1 to reverse of: AAH12192 from: 1 to: 394

682 LysIlePheSerSerAlaThrSer 689
|||||
84 AAAATCTTCTCATCAGCACATCT 61

seq_name: /STD2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH98826

seq_documentation_block:
ID AAH98826 standard; cDNA: 471 BP.
XX
XX AAH98826;
XX
XX 12-OCT-2001 (first entry)
XX

```

```

DE Human EST-derived coding sequence SEQ ID NO: 693.
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
XX MO200154477-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02687.
XX
XX 25-JAN-2000; 2000US-0491404.
XX 17-JUL-2000; 2000US-0617746.
XX 03-AUG-2000; 2000US-0631451.
XX 15-SEP-2000; 2000US-0663870.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi Y,
PI Cao Y, Dzmanac RA, Zhang J, Wehman T;
XX
XX WPI: 2001-476164/51.
XX DR P-PSDB: AAM24167.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
XX Claim 1: Page 629; 1275pp; English.
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
XX Sequence 471 BP; 66 A; 154 C; 122 G; 129 T; 0 other;
XX
XX
XX
XX alignment_scores:
XX      Quality:      8.00      Length:      8
XX      Ratio:      1.000      Gaps:      0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-697-089-2 x AAH98826 ..
XX
XX Align seg 1/1 to: AAH98826 from: 1 to: 471
XX
XX 452 GYARgARgLeuSerLeuLeu 459
XX |||||||||||||||||||
XX 383 GGACGCCGCTGCTCTGCTT 406
XX
XX seq_name: /SID52/gcdata/geneseq/geneseqn/NA2000.DAT: AAC08443
XX
XX seq_documentation_block:
XX ID AAC08443 standard; cDNA; 484 BP.
XX
XX AAC08443:
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 12518.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX

```

```

OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1: SEQ ID 12518; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 484 BP; 105 A; 139 C; 130 G; 108 T; 2 other;
XX
XX
XX
XX alignment_scores:
XX      Quality:      8.00      Length:      8
XX      Ratio:      1.000      Gaps:      0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-697-089-2 x AAC08443/rev ..
XX
XX Align seg 1/1 to reverse of: AAC08443 from: 1 to: 484
XX
XX 645 ProSerARgAlaValSerLeuphe 652
XX |||||||||||||||||||
XX 207 CCGTCCAGGGCGCTGCTCTGCTTC 184
XX
XX seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT: AA110777
XX
XX seq_documentation_block:
XX ID AA110777 standard; DNA; 493 BP.
XX
XX AA110777:
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #710 for gene expression analysis in human cervical cell sample.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX

```

XX 30-JAN-2001; 2001WO-US00670.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 710; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 493 BP; 124 A; 151 C; 117 G; 101 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAI10777/rev ..

Align seg 1/1 to reverse of: AAI10777 from: 1 to: 493

997 LeuGInGluAaArgLeuValGly 1004

|||||

167 CTCGAGGAGCCAGCGTAgTTGGG 144

seq_name: /SID2/gcgcdata/geneseq/geneseqn/NA2001.DAT:AAI32035

seq_documentation_block:

ID AAI32035 standard; DNA; 493 BP.

XX AAI32035;

XX 17-OCT-2001 (first entry)

XX Probe #721 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 721; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 493 BP; 124 A; 151 C; 117 G; 101 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAI32035/rev ..

Align seg 1/1 to reverse of: AAI32035 from: 1 to: 493

997 LeuGInGluAaArgLeuValGly 1004

|||||

167 CTCGAGGAGCCAGCGTAgTTGGG 144

seq_name: /SID2/gcgcdata/geneseq/geneseqn/NA2001.DAT:AAI0702

seq_documentation_block:

ID AAI0702 standard; DNA; 493 BP.

XX AAI0702;

XX 09-OCT-2001 (first entry)

XX Probe #693 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-476286/51.
DR
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 693; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 493 BP; 124 A; 151 C; 117 G; 101 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA100702/rev ..
Align seg 1/1 to reverse of: AA100702 from: 1 to: 493

997 LeuGInGUAlAArgLeuValGly 1004
|||||
167 CTCGAGGAGCCAGCGCTAGTTGGG 144

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AA117293

seq_documentation_block:
ID AA117293 standard; DNA: 509 BP.
XX
AC AA117293;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #7226 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI: 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 25; SEQ ID No 7226; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SNPs). The present sequence is one such probe. The SNPs are derived
CC from human HeLa cells. The SNPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 509 BP; 97 A; 149 C; 132 G; 131 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AA117293 ...
Align seg 1/1 to: AA117293 from: 1 to: 509

316 GluLeuAlaGluGlyLeuLeu 323
|||||
378 GAACCTGCAGAGCGTCTGCTTCTT 401

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AA142181

seq_documentation_block:
ID AA142181 standard; DNA: 509 BP.
XX
AC AA142181;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #10867 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 10867; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 509 BP; 97 A; 149 C; 132 G; 131 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAH12181 ..

Align seg 1/1 to: AAH12181 from: 1 to: 509

316 GLEUENALAGUGLYLEULEU 323
|||||
378 GAACCTGCAGAGGCTCTCTT 401

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH12617

seq_documentation_block:
ID AAH12617 standard; cDNA; 532 BP.
XX
AC AAH12617;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:9452.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
PS Claim 3; SEQ ID 9452; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 532 BP; 172 A; 104 C; 63 G; 183 T; 10 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAH12617 ..

Align seg 1/1 to: AAH12617 from: 1 to: 532

895 Serleuserleuleulyshis 902
|||||
234 TCTTATCAGCCCTCTTAAACAC 257

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX28297

seq_documentation_block:
ID AAX28297 standard; DNA; 591 BP.
XX
AC AAX28297;
XX
DT 17-JUN-1999 (first entry)
XX
DE Human CYP3A4 gene intron 3.
XX
KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;
KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;
KW genetic linkage detection; phenotypic variation; intron; ss.
XX
OS Homo sapiens.
XX
PN WO913106-A1.
XX
PD 18-MAR-1999.
XX
PE 02-SEP-1998; 98WO-US18158.
XX
PR 10-SEP-1997; 97US-0058612.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Guida M, Lichter JB;
XX
DR WPI; 1999-215070/18.
XX
XX
PT New isolated CYP3A4 polymorphic sequences
XX
PS Disclosure; Page 29; 40pp; English.
XX
XX
CC This sequence represents an intron of the human CYP3A4 gene.
CC The invention relates to a CYP3A4 sequence polymorphism,
CC which is part of a non-naturally occurring chromosome. Nucleic acids

CC comprising the CYP3A4 polymorphic sequences can be used to screen
CC patients for altered metabolism for CYP3A4 substrates, potential
CC drug-drug interactions, and adverse/side effects as well as diseases that
CC result from environmental or occupational exposure to toxins. They can
CC also be used to establish animal, cell culture and in vitro cell-free
CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used
CC for expression studies to determine the effect of promoter and/or intron
CC sequence variations on mRNA expression and stability. The polymorphisms
CC are also used as single nucleotide polymorphisms to detect genetic
CC linkage to phenotypic variation in activity and expression of CYP3A4. The
CC nucleic acids can also be used to generate genetically modified non-human
CC animals or site specific gene modifications in cell lines.

XX
SQ Sequence 591 BP; 177 A; 106 C; 134 G; 174 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAX28297/rev ..

Align seg 1/1 to reverse of: AAX28297 from: 1 to: 591

361 ThrLeuphenisthrPheTyrAsp 368
|||||
49 ACACCTTTTCATCTTTTATGAC 26

seq_name: /SIDB2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC95560

seq_documentation_block:

ID AAC95560 standard; cDNA; 643 BP.

XX AAC95560;

AC AAG95560;

DT 21-FEB-2001 (first entry)

DE Human secreted protein gene 40 SEQ ID NO:50.

XX Human: secreted protein; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiac; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease;
KW ss.

XX Homo sapiens.

XX WO200061596-A1.

XX 19-OCT-2000.

PD 06-APR-2000; 2000WO-US08983.

PF 09-APR-1999; 99US-0128703.

PR 20-JAN-2000; 2000US-0176068.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Ruben SM, Komatsoulis G;

DR WPI: 2000-611865/58.

DR P-PSDB: AAB52051.

XX Fifty nucleic acid molecules encoding human secreted proteins, useful
PT in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases -
XX
PS Claim 1, Page 440; 505pp; English.

CC polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
CC AAB52103 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnary;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiac.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention.

XX
SQ Sequence 643 BP; 105 A; 184 C; 211 G; 143 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAC95560 ..

Align seg 1/1 to: AAC95560 from: 1 to: 643

891 AspValGlnGlySerLeuSerSer 898
|||||
97 GATGTACAGGCGAGTCTCTTCC 120

seq_name: /SIDB2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF14778

seq_documentation_block:

ID AAF14778 standard; cDNA; 644 BP.

XX AAF14778;

AC AAF14778;

DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:7301.

XX Multiple gene expression; filamentous fungal cell; EST;

XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;

XX Aspergillus oryzae; Trichoderma reesei; Identification; recombination;

XX culture condition; environmental stress; spore morphogenesis;

XX metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus oryzae.

XX WO200056762-A2.

XX 28-SEP-2000.

PD 22-MAR-2000; 2000WO-US07781.

PF 22-MAR-1999; 99US-0273623.

PR (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

DR WPI: 2000-594572/56.
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX
PS Claim 88: Page 2952; 3161pp; English.
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SO Sequence 644 BP: 190 A; 154 C; 158 G; 141 T; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAF14778/rev ..

Align seg 1/1 to reverse of: AAF14778 from: 1 to: 644

415 SerValAsnGluAspValLeuLeu 422

|||||

252 AGTGTGAACGAGATGTTCTATTG 229

seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH42838

seq_documentation_block:

ID AAH42838 standard; cDNA: 650 BP.

XX AC AAH42838;

XX DT 01-OCT-2001 (first entry)

XX DE Nucleotide sequence of a human trans-glog1 p230 gene.

XX KM Cancer associated gene; prostate cancer; stomach cancer;

XX KM oesophageal cancer; cancer; vaccine; ss.

OS Homo sapiens.

XX PN WO200153524-A2.

XX PD 26-JUL-2001.

XX PF 18-JAN-2001; 2001WO-GB00188.

XX PR 18-JAN-2000; 2000GB-0000993.

XX
PA (UYNO-) UNITV NOTTINGHAM TRENT.
XX
XX
PI Rees RC, Li G, Mian S;
XX
XX
DR WPI: 2001-476121/51.
XX
XX
PT Use of cancer-associated genes and their products in detecting,
PT monitoring, treating or preventing cancer, specifically prostate
PT cancer, and in developing DNA-based vaccines that promote anti-tumor
PT responses -
XX
PS Claim 1: Page 29-30; 43pp; English.
XX
XX AAH42781-AAH42846 represent the nucleotide sequences of cancer
CC associated genes, identified using SEREX (Serological Identification of
CC Antigens by Recombinant Expression Cloning). The genes are overexpressed
CC in prostate cancer, and some are overexpressed in other cancers such as
CC stomach cancer and oesophageal cancer. The nucleic acids are useful for
CC detecting and monitoring cancer, particularly prostate cancer. They are
CC particularly useful in the treatment or prevention of cancer. In
CC producing DNA-based vaccines against prostate cancer or that promote
CC anti-tumor immune responses, and to raise antibodies. The expression of
CC genes and detection of their protein products and/or peptides may be
CC used to monitor disease progression during therapy, or as a prognostic
CC indicator of initial disease status of the patient.
XX
SO Sequence 650 BP: 165 A; 158 C; 171 G; 139 T; 17 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAH42838/rev ..

Align seg 1/1 to reverse of: AAH42838 from: 1 to: 650

415 SerValAsnGluAspValLeuLeu 422

|||||

355 TCTGTAAATGAGATGTCCTGCTC 332

seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH07394

seq_documentation_block:

ID AAH07394 standard; cDNA: 656 BP.

XX AC AAH07394;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA clone (5'-primer) SEQ ID NO:4229.

XX KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

93 TTTGCTTTTCTCCGCTCTCT 116

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AAH07749

seq_documentation_block:

ID AAH07749 standard; cDNA: 837 BP.

AAH07749:

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:4584.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI: 2001-318749/34.

Claim 1; SEQ ID 4584; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH2446 to AAH93893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 837 BP; 259 A; 135 C; 217 G; 220 T; 6 other:

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAH07749/rev ..

Align seg 1/1 to reverse of: AAH07749 from: 1 to: 837

453 ArgArgLeuSerSerLeuLeuThr 460

|||||
264 CCGAGACTGTCTCTCTGCTAACCA 241

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AAH046452

seq_documentation_block:

ID AAH046452 standard; DNA: 855 BP.

AAH046452:

18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 50193.

Hybridisation assay; genetic mapping; gene expression control;

protein identification; signal transduction pathway;

metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

06-MAY-1999; 99US-0132487.

07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139465.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151303.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161306.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 8.00
Ratio: 1.000
Percent Similarity: 100.000
Length: 8
Gaps: 0
Percent Identity: 100.000

```
alignment_block:
US-09-697-089-2 x AAC46452 ..
Align seg 1/1 to: AAC46452 from: 1 to: 855
703 SerleuSerleuValleuSerThr 710
|||||
793 TCCCTTAGTTGTGCTCTCGACC 816

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC41255
seq_documentation_block:
ID AAC41255 standard; DNA: 955 BP.
XX
AC AAC41255;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 31218.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
```

```
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145115.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
```


PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157863.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 8 00
Ratio: 1.000
Percent Similarity: 100.000
Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAC41255 ..
Align seg 1/1 to: AAC41255 from: 1 to: 955
454 ArgLeuSerSerLeuLeuThrSer 461
|||||
356 CGATTGTCGAGTCTCTTAACATCG 379

seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAC47554

seq_documentation_block:

ID AAC47554 standard; DNA; 1076 BP.

AC AAC47554;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54251.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x MAC47554/rev ..

Align seg 1/1 to reverse of: AAC47554 from: 1 to: 1076

986 LysleuserGlnValleuserLys 993

|||||
640 AACCTCACTCAAGTACTCTCTAAG 617

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC35424

seq_documentation_block:

ID AAC35424 standard; DNA: 1078 BP.

XX AAC35424;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10144.

XX Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135253.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

```

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:
 Quality: 8.00
 Ratio: 1.000
 Percent Similarity: 100.000
 Length: 8
 Gaps: 0
 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AAC35424/rev

```

Align seg 1/1 to reverse of: AAC35424 from: 1 to: 1078
986 LysleuSerGlnValLeuSerLys 993
|||||
640 AACTCAGTCAGTACTGTCTTANG 617

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA47543
seq_documentation_block:
ID AAA47543 standard; DNA: 1089 BP.
XX
AC AAA47543;
XX
DT 20-OCT-2000 (first entry)
XX
DE Primate GPR2 sequence.
XX
KW Cutaneous T-cell attracting chemokine; CTACK; skin; cell movement;
KW migration; vasoactive intestinal contractor; Vic; GPR2; agonist;
KW antagonist; antibody; immunological condition; mutein; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1089
FT /*tag= a
FT /*product= GPR2
XX
PM WO200038713-A1.
XX
PD 06-JUL-2000.
XX
PF 23-DEC-1999; 99WO-US30819.
XX
PR 24-DEC-1998; 98US-0113858.
PR 27-MAY-1999; 99US-0322580.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Wang W, Oldham ER, Soto H, Lui Y, Hudak SA, Homey B, Morales JM,
PI Kellermann S, McEvoy LM, Zlotnik A;
XX
DR WPI: 2000-465633/40.
DR P-PSDB: AAB01448.
XX
PT Modulating cell movement within the skin, useful for treating
PT immunological skin conditions or diseases comprises administering T
PT cell-attracting chemokine or vasoactive intestinal contractor chemokine
PT agonists or antagonists
XX
PS Example 3; Page 62-64; 79pp; English.
XX
CC Modulating movement of a cell within or to the skin of a mammal can
CC be achieved by administering an antagonist or agonist of cutaneous T
CC cell-attracting chemokine (CTACK) or vasoactive intestinal contractor
CC (Vic) chemokine. The antagonist is selected from a mutein of natural
CC CTACK or Vic, an antibody which neutralises CTACK or Vic or an
CC antibody which block GPR2 ligand binding. The CTACK or Vic agonists
CC or antagonists are useful for treating medical conditions or diseases
CC associated with immunological conditions of the skin.
XX
SQ Sequence 1089 BP; 127 A; 409 C; 347 G; 206 T; 0 other;

```

alignment_scores:
 Quality: 8.00
 Ratio: 1.000
 Percent Similarity: 100.000
 Length: 8
 Gaps: 0
 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AAA47543/rev

Align seg 1/1 to reverse of: AAA47543 from: 1 to: 1089

685 SerSerAlaThrSerLeuArgLeu 692
|||||
878 AGCAGTGCAGACATCCTGCGTTTG 855

seq_name: /STDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT.AAC33764

seq_documentation_block:

ID AAC33764 standard; DNA; 1133 BP.

XX AAC33764;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 4232.

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139753.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157863.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-697-089-2 x AAC33764/rev ..

Align seg 1/1 to reverse of: AAC33764 from: 1 to: 1133

893 GlnGlySerLeuSerLeuLeu 900
|||||
290 CAAAGCTCACTACTCTCCTGCTG 267
seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC51511
seq_documentation_block:
ID AAC51511 standard; DNA: 1133 BP.
XX
AC AAC51511;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68798.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142955.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-697-089-2 x AAC51511/rev ..

Align seq 1/1 to reverse of: AAC51511 from: 1 to: 1133

893 GInGlySerLeuSerLeuLeu 900

```
|||||
290 CAAGGCTCACTATCTCTGGCG 267

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA1993.DAT:AAQ46074
seq_documentation_block:
ID AAQ46074 standard; DNA: 1152 BP.
XX
AC AAQ46074;
XX
DT 08-FEB-1994 (first entry)
XX
DE Partial sequence of the gene ERG12 which incodes mevalonate kinase
DE (MK).
XX
KM Mevalonate kinase; enyme; phylosteroel biosynthesis; plant hormone; ss.
XX
OS Yeast.
XX
FH Key Location/Qualifiers
FT CDS 580..1152
FT /*tag= a
FT misc_feature 577..603
FT /*tag= b
FT /*note= "ON1"
FT misc_feature 1009..1026
FT /*tag= c
FT /*note= "ON2"
XX
PN WO9316187-A.
XX
PD 19-AUG-1993.
XX
PF 09-FEB-1993; 93WO-FR00134.
XX
PR 14-FEB-1992; 92FR-0001712.
XX
PA (VERN-) VERNOUTIL RECH.
XX
PI Karst F, Lejeune F, Oulmouden A, Tourte M;
XX
DR WPI: 1993-272890/34.
DR P-PSDB: AAR39562.
XX
PT Plants modified with genes encoding enzymes of phylosteroel
PT biosynthesis - esp. mevalonate kinase, show improved regeneration
PT capacity and better productivity
XX
PS Disclosure; Figure 14; 75pp; French.
XX
CC The inventors claim vector pFAB2 (CNCM 1-1176) contg. ERG12 from
CC yeast and the cauliflower mosaic virus promoter.
CC This plasmid was inserted, by conjugation, into A.tumefaciens strain
CC LBA 4404, in presence of pRK2013 as helper plasmid. Cells able to
CC grow on medium contg. kanamycin, streptomycin and rifampicin were
CC selected and used to transform tobacco leaf discs. The resulting
CC plants were grown to flowering and fragments of the leaves
CC regenerated. Shoots were visible after 3 weeks (without any evident
CC callus formation); these grew rapidly and additional shoots
CC developed. Complete plants, suitable for transfer to soil, were
CC produced within about 4 months. NB, AAQ46074 does not translate into
CC the AA sequence listed in Figure 14.
XX
SQ Sequence 1152 BP; 297 A; 9 C; 473 G; 373 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAQ46074/rev
```

```
Align seg 1/1 to reverse of: AAQ46074 from: 1 to: 1152
354 PHEHSSERHISTHRCGLNHRTHR 361
|||||
442 TTTCACCTCCACACCCACACACC 419

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA1999.DAT:AAx27354
seq_documentation_block:
ID AAX27354 standard; DNA: 1220 BP.
XX
AC AAX27354;
XX
DT 11-JUN-1999 (first entry)
XX
DE Human secreted protein gene 44 clone HHDAD22.
XX
KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9902546-A1.
XX
PD 21-JAN-1999.
XX
PF 07-JUL-1998; 98WO-US13684.
XX
PR 12-SEP-1997; 97US-0058785.
PR 08-JUL-1997; 97US-0051916.
PR 08-JUL-1997; 97US-0051918.
PR 08-JUL-1997; 97US-0051919.
PR 08-JUL-1997; 97US-0051920.
PR 08-JUL-1997; 97US-0051925.
PR 08-JUL-1997; 97US-0051926.
PR 08-JUL-1997; 97US-0051928.
PR 08-JUL-1997; 97US-0051929.
PR 08-JUL-1997; 97US-0051930.
PR 08-JUL-1997; 97US-0051931.
PR 08-JUL-1997; 97US-0051932.
PR 08-JUL-1997; 97US-0052732.
PR 08-JUL-1997; 97US-0052733.
PR 08-JUL-1997; 97US-0052793.
PR 08-JUL-1997; 97US-0052795.
PR 08-JUL-1997; 97US-0052803.
PR 18-AUG-1997; 97US-0055684.
PR 18-AUG-1997; 97US-0055722.
PR 18-AUG-1997; 97US-0055723.
PR 18-AUG-1997; 97US-0055947.
PR 18-AUG-1997; 97US-0055948.
PR 18-AUG-1997; 97US-0055949.
PR 18-AUG-1997; 97US-0055950.
PR 18-AUG-1997; 97US-0055953.
PR 18-AUG-1997; 97US-0055954.
PR 18-AUG-1997; 97US-0055964.
PR 18-AUG-1997; 97US-0055984.
PR 18-AUG-1997; 97US-0056360.
PR 12-SEP-1997; 97US-0058660.
PR 12-SEP-1997; 97US-0058661.
PR 12-SEP-1997; 97US-0058664.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PA Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA;
XX
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;
```


DR WPI; 1999-120770/10.
 DR P-PSDB; AAY02693.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1: Page 272; 464pp: English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAX27302) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 123 novel genes and their fragments (nucleic
 CC acid sequences: AAX27311-X27449; amino acid sequences AAY02650-102788)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 123
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX27311 for described uses).
 CC
 SQ Sequence 1220 BP; 335 A; 214 C; 242 G; 426 T; 3 other:
 XX
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x AAX27354 ..
 Align seg 1/1 to: AAX27354 from: 1 to: 1220
 199 PheValPhePheLeuArgLeuSer 206
 ||||||||||||||||||
 253 TTTGTTTTCCTTCGACACTGACT 276
 seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA1996.DAT:AA733325
 seq_documentation_block:
 ID AA733325 standard: DNA; 1225 BP.
 XX
 AC AA733325;
 XX
 DT 18-OCT-1996 (first entry)
 XX
 DE American elm chitinase-like protein coding sequence.
 XX
 KW Clone PHS2; chitinase; american elm; fungus; Ophiostoma ulmi;
 KW dutch elm disease; E. coli; chitinase-like protein;
 KW fungal infection; ss.
 XX
 OS American elm.
 XX
 FH Key
 FH sig_peptide Location/Qualifiers
 FT 88..150
 FT /*tag= a
 FT mat_peptide 151..1038
 FT /*tag= b
 FT CDS 88..1041
 FT /*tag= c
 FT /product= Chitinase-like protein
 FT 1147..1151
 FT /*tag= d
 FT polyA_signal 1153..1158
 FT /*tag= e
 FT polyA_signal 1168..1177
 FT /*tag= f

XX
 PN US5539095-A.
 XX
 PD 23-JUL-1996.
 XX
 PF 04-AUG-1994; 94US-0286020.
 XX
 PR 04-AUG-1994; 94US-0286020.
 XX
 PA (UNMS) UNIV MICHIGAN STATE.
 XX
 PI Hajela RK, Sticklen MB;
 XX
 DR WPI; 1996-353879/35.
 DR P-PSDB; AAM00186.
 XX
 PT Isolated chitinase gene derived from an American elm - used to
 PT obtain prods. for inhibiting fungal infection of plants
 XX
 PS Claim 1: Fig 1; 12pp: English.
 XX
 CC This sequence represents the cDNA clone PHS2, which encodes a protein
 CC having chitinase-actively derived from american elm. This protein
 CC inhibits the fungus Ophiostoma ulmi, the causative agent of dutch
 CC elm disease. The clone PHS2 may be used to transform E. coli cells
 CC for the recombinant production of the chitinase-like protein. The
 CC protein may be used in a composition to inhibit fungal infection of
 CC elm trees.
 CC
 SQ Sequence 1225 BP; 322 A; 277 C; 313 G; 313 T; 0 other:
 XX
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x AA733325 ..
 Align seg 1/1 to: AA733325 from: 1 to: 1225
 452 GlyArgArgLeuSerLeuLeu 459
 ||||||||||||||||||
 1085 GGCCGTCGTTTCGTCACACTTCTT 1108
 seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA1996.DAT:AA729171
 seq_documentation_block:
 ID AA729171 standard: DNA; 1263 BP.
 XX
 AC AA729171;
 XX
 DT 24-JUL-1996 (first entry)
 XX
 DE Adenovirus 11 subtype B (Ad-11) fragment, shows residues 1645-2907.
 XX
 KW Polymerase chain reaction; PCR; amplify; primer; probe; detection;
 KW identification; adenovirus; exon region; serotype; subtype; ds.
 XX
 OS Mastadenovirus.
 XX
 PN JP07327700-A.
 PD 19-DEC-1995.
 XX
 PF 08-JUN-1994; 94JP-0126163.
 XX
 PR 08-JUN-1994; 94JP-0126163.
 XX
 PA (MITP) MITSUBISHI YUKA BCL KK.
 XX
 DR WPI; 1996-072347/08.

DR P-PSDB; AAR97600.
XX
PT Detection and identification of adenovirus using serotype and
PT sub-type specific oligo:nucleotide(s) - and probes and primers used
PT in the method
XX
PS Claim 40; Page 32-33; 42pp; Japanese.
XX
CC The sequences given in AAT29165-74 represents exon fragments derived
CC from different subtypes of adenovirus using the primer and probe
CC sequences given in AAT10167-80. These primers and probes are used in
CC the detection and identification of adenovirus. An adenoviral exon
CC region is amplified which has a serotype and subtype specific sequence
CC and then the amplified fragment is detected. The primers used for the
CC amplification are complementary to sequence which are serotype specific.
CC The method allows reliable, rapid and easy detection and identification
CC of the subtype and serotype of adenovirus.
XX
SQ Sequence 1263 BP; 314 A; 362 C; 251 G; 318 T; 18 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAT29171/rev ..

Align seg 1/1 to reverse of: AAT29171 from: 1 to: 1263

729 GIUARGHISIEThrServaIThr 736
|||||
78 GAAGGACATTAACGTCCTTACC 55

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT29172
seq_documentation_block:
ID AAT29172 standard; DNA; 1263 BP.
XX
AC AAT29172;
XX
DT 24-JUL-1996 (first entry)
XX
DE Adenovirus 14 subtype B (Ad-14) fragment, shows residues 1645-2907.
XX
KW Polymerase chain reaction; PCR; amplify; primer; probe; detection;
KW identification; adenovirus; exon region; serotype; subtype; ds.
XX
OS Mastadenovirus.
XX
XX
XX JP07327700-A.
XX
PD 19-DEC-1995.
XX
PF 08-JUN-1994; 94JP-0126163.
XX
PR 08-JUN-1994; 94JP-0126163.
XX
PA (MITP) MITSUBISHI YUKA BCL KK.
XX
XX WPI; 1996-072347/08.
XX
DR P-PSDB; AAR97601.
XX
PT Detection and identification of adenovirus using serotype and
PT sub-type specific oligo:nucleotide(s) - and probes and primers used
PT in the method
XX
PS Claim 41; Page 33-35; 42pp; Japanese.
XX
CC The sequences given in AAT29165-74 represents exon fragments derived
CC from different subtypes of adenovirus using the primer and probe
CC sequences given in AAT10167-80. These primers and probes are used in

CC the detection and identification of adenovirus. An adenoviral exon
CC region is amplified which has a serotype and subtype specific sequence
CC and then the amplified fragment is detected. The primers used for the
CC amplification are complementary to sequence which are serotype specific.
CC The method allows reliable, rapid and easy detection and identification
CC of the subtype and serotype of adenovirus.
XX
SQ Sequence 1263 BP; 317 A; 355 C; 254 G; 329 T; 8 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAT29172/rev ..

Align seg 1/1 to reverse of: AAT29172 from: 1 to: 1263

729 GIUARGHISIEThrServaIThr 736
|||||
78 GAAGGACATTAACGTCCTTACC 55

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD03929
seq_documentation_block:
ID AAD03929 standard; cDNA; 1322 BP.
XX
AC AAD03929;
XX
DT 02-JUL-2001 (first entry)
XX
DE Soybean NADH oxidase-like AIF cgs cDNA contig.
XX
KW Soybean; NADH oxidase-like AIF; apoptosis inducing factor;
KW transgenic plant; male sterility; plant architecture alteration;
KW plant breeding; ss.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT CDS 68..1153
FT
FT /*tag= a
FT /product= "Soybean NADH-oxidase-like AIF"
FT /transl_except= (pos:338..340, aa:Xaa)
FT /note= "Xaa is an unknown amino acid"
XX
XX W0200125270-A2.
XX
XX
XX 12-APR-2001.
XX
PD 27-SEP-2000; 2000WC-US26441.
XX
PF 01-OCT-1999; 990S-0157311.
XX
PR (DUPC) DU PONT DE NEMOURS & CO E I.
XX
PA Cahoon RE, Mazithulela G;
XX
XX WPI; 2001-290606/30.
XX
DR P-PSDB; AAE00631.
XX
PT New nucleic acids encoding apoptosis inducing factors in plants and
PT seeds, useful in plant breeding approaches, for developing lines with
PT desired phenotypes -
XX
XX
XX Claim 3; Page 48; 56pp; English.
XX
CC The present sequence is a cgs cDNA encoding soybean NADH oxidase-like
CC apoptosis inducing factor (AIF) of the invention. This sequence is
CC a contig of clones s11.pk0105.f2, s12.pk125.g24, s11.pk0029.e11,
CC s11.pk0078.a10, s11.pk0128.c10 and s113c.pk003.c7. NADH oxidase-like

CC AIF sequences are used to alter apoptosis in plant cells and to control
 CC cell tissue culture growth. The polynucleotides or nucleic acid fragments
 CC of the invention are useful for isolating cDNAs and genes encoding
 CC homologous proteins from the same or other plant species. They are
 CC particularly useful in plant breeding approaches, such as for developing
 CC lines with desired phenotypes. They are also used to create transgenic
 CC plants in which the NADH oxidase-like protein are present at higher or
 CC lower levels than normal or in cell types or developmental stages in
 CC which they are not normally found. This has the effect of altering
 CC apoptosis and programmed cell death in those cells. Tissue-specific
 CC expression of AIF is used to selectively kill defined cell types and
 CC therefore influence developmental pathways in plant organs which is
 CC useful for male sterility, altering plant architecture, growth of cells
 CC and tissue in culture and for improving transformation by controlling
 CC death caused by the trauma of particle bombardment.

XX
 SQ Sequence 1322 BP; 434 A; 216 C; 309 G; 362 T; 1 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AAD03929/rev ..

Align seg 1/1 to reverse of: AAD03929 from: 1 to: 1322

595 LeuPheasphephegluH1sleu 602
 ||||||||||||||||||
 889 TTGTTGATTTCTTTGAACATCTG 866

seq_name: /STD52/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC76711

seq_documentation_block:
 ID AAC76711 standard; cDNA; 1414 BP.

XX
 AC AAC76711;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF2266 polynucleotide sequence SEQ ID NO:4531.

XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnertary; antipaporiatic; antiparkinsonian; noctropic; neuroprotective;
 KW anticonvulsant; osteopathic; antialthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX
 OS Homo sapiens.
 PN MO200058473-A2.
 PD 05-OCT-2000.
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.

XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB: AAB42502.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

XX
 PS Claim 5; Page 3729; 5507pp; English.

XX
 AC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnertary;
 CC antipaporiatic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antialthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypoid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX
 SQ Sequence 1414 BP; 394 A; 302 C; 344 G; 372 T; 2 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AAC76711/rev ..

Align seg 1/1 to reverse of: AAC76711 from: 1 to: 1414

689 SerLeuArgLeuGlnIleLysArg 696
 ||||||||||||||||||
 1373 TCCCTGAGGCTACGATTAGAGG 1350

seq_name: /STD52/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC49484

seq_documentation_block:
 ID AAC49484 standard; DNA; 1508 BP.

XX
 AC AAC49484;
 DT 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 61329.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW Protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

XX
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 PD 06-SEP-2000.
 PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 03-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126254.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1993; 99US-0156596.
PR 04-OCT-1993; 99US-0157117.
PR 05-OCT-1993; 99US-0157753.
PR 06-OCT-1993; 99US-0157865.
PR 07-OCT-1993; 99US-0158029.
PR 08-OCT-1993; 99US-0158232.
PR 12-OCT-1993; 99US-0158369.
PR 13-OCT-1993; 99US-0159293.
PR 13-OCT-1993; 99US-0159294.
PR 13-OCT-1993; 99US-0159295.
PR 14-OCT-1993; 99US-0159329.
PR 14-OCT-1993; 99US-0159330.
PR 14-OCT-1993; 99US-0159331.
PR 14-OCT-1993; 99US-0159637.
PR 14-OCT-1993; 99US-0159638.
PR 18-OCT-1993; 99US-0159584.
PR 21-OCT-1993; 99US-0160741.
PR 21-OCT-1993; 99US-0160767.
PR 21-OCT-1993; 99US-0160770.
PR 21-OCT-1993; 99US-0160814.
PR 21-OCT-1993; 99US-0160815.
PR 22-OCT-1993; 99US-0160980.
PR 22-OCT-1993; 99US-0160981.
PR 22-OCT-1993; 99US-0160989.
PR 25-OCT-1993; 99US-0161404.
PR 25-OCT-1993; 99US-0161405.
PR 25-OCT-1993; 99US-0161406.
PR 26-OCT-1993; 99US-0161359.
PR 26-OCT-1993; 99US-0161360.
PR 26-OCT-1993; 99US-0161361.
PR 28-OCT-1993; 99US-0161920.
PR 28-OCT-1993; 99US-0161992.
PR 28-OCT-1993; 99US-0161993.
PR 29-OCT-1993; 99US-0162142.

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAC49484 ..

Align seg 1/1 to: AAC49484 from: 1 to: 1508

320 GlycylleuleuglnIleGlnlys 327

1225 GGTGTTGTTACTTCMAATCCAGAAA 1248

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT:AA087717

seq_documentation_block:

ID AA087717 standard; cDNA; 1512 BP.

AC AA087717;

DT 14-NOV-1995 (first entry)

DE Human cytochrome P450 molecular species 3A4 cDNA coding region.

KW Human cytochrome P450; amplification; PCR; primer; expression vector;

KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;

KW Carcinogen; mutagen; liver metabolism; ds.

OS Homo sapiens.

XX EP644267-A.

XX 22-MAR-1995.

PF 20-JUL-1994; 94EP-0111298.

XX

PR 21-JUL-1993; 93UP-0180246.
PR 20-JUL-1993; 93UP-0201120.
PR 30-JUL-1993; 93UP-0208279.
XX
PA (HAYASHI) HAYASHI K.
PA (SUMO) SUMITOMO CHEM CO LTD.
PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
PI Yabusaki Y;
XX
DR WPI: 1995-116991/16.
DR P-PSDB; AAR72363.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
PS Examples; Page 31-33; 124pp; English.
XX
CC The nucleotide sequence of the cDNA coding region for the human
CC cytochrome P450 species 3A4. The gene encodes a protein of 503 amino
CC acids. The cDNA was amplified by PCR using the primers AA087743-6. The
CC product was cloned into the yeast expression vectors pAAH5N or pAHR to
CC produce the vectors p3A4 for the expression of the cytochrome P450 alone
CC or p3A4R for co-expression with the yeast NADPH-P450 reductase.
CC The vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (AA087714), 2C9 (AA087715),
CC 2E1 (AA087716), or 3A4 or their auxiliary species and variants
CC (AA087718-32), and yeast NADPH-P450 reductase, either as a fused protein
CC or in cell extracts, and analysing the resulting metabolite to assess the
CC safety of the chemical compound. The method is useful for determining
CC whether the chemical compound, or its metabolite, will be converted into
CC a carcinogenic or mutagenic form through metabolism in the liver.
XX
SQ Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 other;

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AA087717/rev ..

Align seg 1/1 to reverse of: AA087717 from: 1 to: 1512

361 ThrleuphehistrPhetyrasp 368

214 ACACCTTTCATCTACTTTTATGAC 191

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT28383

seq_documentation_block:

ID AAT28383 standard; DNA; 1512 BP.

AC AAT28383;

DT 11-OCT-1996 (first entry)

DE Human cytochrome P450 molecular species 3A4 gene.

KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;

KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;

KW evaluation; safety; fusion protein; metabolite; detoxification;

KW carcinogenic; ds.

OS Homo sapiens.

XX JP08056695-A.

PD 05-MAR-1996.

XX

PF 15-JUL-1994; 94JP-0164184.
XX
PR 17-JUN-1994; 94JP-0136053.
PR 20-JUL-1993; 93JP-0201120.
PR 30-JUL-1993; 93JP-0208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX WPI: 1996-182311/19.
DR P-PSDB: AAR93170.
XX
PT Novel method for the evaluation of the safety of a cpd. - using a
PT human cytochrome P450 and yeast NADPH reductase to determine whether
PT the analyte cpd. is detoxified or metabolised to a carcinogen
XX
PS Example 1; Page 24-26; 74pp; Japanese.
XX
CC This is the nucleotide sequence of the human cytochrome P450 molecular
CC species 3A4 gene which encodes a protein of 503 amino acids. The gene
CC was amplified from a human liver derived cDNA library as 2 fragments of
CC 0.6 and 0.9 kb using primers AAT26933-6. The prod. was cloned into the
CC yeast expression vector pAH5N to generate plasmid p3A4 for prodn. of
CC the cytochrome only or into the vector pAHRR to generate the plasmid
CC p3A4R for co-prodn. with the yeast NADPH-P450 reductase. The sequence
CC is placed under control of the yeast ADH gene promoter and terminator.
CC The vectors are used in a method for evaluating the safety of a cpd. by
CC reacting the test cpd. with recombinantly produced human cytochrome P450
CC mol. species 1A2 (AAT28380), 2C9 (AAT28381), 2E1 (AAT28382), 3A4 or
CC their variants (AAT28384-98) together with yeast NADPH-P450 reductase
CC (either as a fused protein or as a cell extract) and analysing the
CC resultant metabolite. The cpd. is considered "safe" if it is detoxified
CC or not rendered carcinogenic or "unsafe" if it is not detoxified or is
CC metabolised to a carcinogenic cpd.
SQ Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAT28383/rev ..

Align seg 1/1 to reverse of: AAT28383 from: 1 to: 1512

361 ThrLeuPheHisThrPheTyrAsp 368
|||||
214 ACACCTTTCCATACCTTTTATGAC 191

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA1996.DAT:AAT17399

seq_documentation_block:
ID AAT17399 standard; cDNA; 1512 BP.
XX
AC AAT17399;
XX
DT 01-AUG-1996 (first entry)
XX
DE Human derived cytochrome P4503A4 cDNA.
XX
KW Human derived cytochrome; P4503A4; commercial cDNA library; yeast;
KW transfection; recombinant production; expression vector; mammal;
KW immunisation; sensitisation; antibody; determination; detection;
KW non-cross reactive; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1512
FT /*tag= a
XX
PN JF08027197-A.

XX
PD 30-JAN-1996.
XX
XX
PF 13-JUL-1994; 94JP-0161552.
XX
PR 13-JUL-1994; 94JP-0161552.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX WPI: 1996-136338/14.
DR P-PSDB: AAR81464.
XX
PT Antibody recognising human derived cytochrome P4503A4 - allows
PT specific detection of cytochrome P450 species in humans
XX
PS Example 1; Pages 10-12; 13pp; Japanese.
XX
CC The present sequence encodes the human derived cytochrome (HDC)
CC P4503A4, which was obtd. from a commercial cDNA library. Yeast
CC were transfected with an expression vector contg. the HDC cDNA,
CC cultured and then disrupted to give a microsomal fraction. The
CC HDC was purified from the fraction, and used to immunise and
CC sensitise a mammal. Blood was drawn from the mammal, and an
CC anti-HDC antibody isolated. The antibody obtd. recognises HDC
CC P4503A4, partic. at a serum dilution rate of 1:10000, and is
CC substantially without cross reaction to other HDC P450 spp..
SQ Sequence 1512 BP; 441 A; 330 C; 330 G; 411 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAT17399/rev ..

Align seg 1/1 to reverse of: AAT17399 from: 1 to: 1512

361 ThrLeuPheHisThrPheTyrAsp 368
|||||
214 ACACCTTTCCATACCTTTTATGAC 191

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AAC32893

seq_documentation_block:
ID AAC32893 standard; DNA; 1513 BP.
XX
AC AAC32893;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 1029.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000
```

```
alignment_block:
US-09-697-089-2 x AAC32893 ..
```

```
Align seg 1/1 to: AAC32893 from: 1 to: 1513
```

```
320 GYLEULEULEGn11eglnly 327
|||||
1226 GGTTTGTTACTCAATCCAGAAA 1249
```

```
seq_name: /SID52/gcdata/geneseq/NA2000.DAT:AAF21023
```

```
seq_documentation_block:
```

```
ID AAF21023 standard; DNA: 1569 BP.
```

```
AC AAF21023;
```

```
DT 14-MAR-2001 (first entry)
```

```
XX Human low adenosine antisense oligonucleotide related sequence #2590.
```

```
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antihasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasocostriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
```

```
XX Homo sapiens.
```

```
XX WO200062736-A2.
```

```
XX 26-OCT-2000.
```

```
XX 24-MAR-2000; 2000WO-US08020.
```

```
XX
```

```
PR 06-APR-1999; 99US-0127958.
XX (UVEC-) UNIV EAST CAROLINA.
PA (NCEC/) NYCE J W.
XX NYCE JW;
XX WPI; 2000-679539/66.
```

```
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions
```

```
PS Disclosure; Page 833; 1592pp; English.
```

```
XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antihasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
```

```
XX Sequence 1569 BP; 440 A; 366 C; 342 G; 421 T; 0 other;
```

```
alignment_scores:
```

```
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000
```

```
alignment_block:
US-09-697-089-2 x AAF21023 ..
```

```
Align seg 1/1 to: AAF21023 from: 1 to: 1569
```

```
700 ValAlaGlySerLeuSerLeuVal 707
|||||
185 GTGGCTGCATCCCTCAGTTGGTG 208
```

```
seq_name: /SID52/gcdata/geneseq/NA2000.DAT:AAA34901
```

```
seq_documentation_block:
```

```
ID AAA34901 standard; DNA: 1569 BP.
```

```
AC AAA34901;
```

```
DT 28-JUL-2000 (first entry)
```

```
XX Human adenosine receptor related polynucleotide SEQ ID NO:2590.
```

```
XX
```


KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW.
XX
DR WPI: 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension, or
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 762-763; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytosolic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32313 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 1569 BP; 440 A; 366 C; 342 G; 421 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAA34901 ..
Align seg 1/1 to: AAA34901 from: 1 to: 1569
700 ValAAGlySerLeuSerLeuVal 707
|||||

185 GTGGCTGGATCCTCAGTTGGTG 208
seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC46641
seq_documentation_block: /
ID AAC46641 standard; DNA: 1655 BP.
XX
AC AAC46641;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 50882.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.

```
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144682.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157717.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

alignment_scores: 8.00
Quality: 1.000
Ratio: 1.000
Percent Similarity: 100.000
Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAC46641

Align seg 1/1 to: AAC46641 from: 1 to: 1655

453 ArgargleuserSerleuethr 460
|||||
1326 AGGAGCGTCGATGCTGCTGACG 1349

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAZ96314
seq_documentation_block:
ID   AAZ96314 standard: DNA; 1729 BP.
XX
XX   AAZ96314;
AC
XX
XX   10-APR-2000 (first entry)
DT
XX
DE   S. pneumoniae derived DNA from ORF #142.
XX
XX   Treatment; prevention; disease; diagnosis; gene therapy; screening;
KM   bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS   Streptococcus pneumoniae.
XX
XX   WO9806734-A1.
PN
XX
XX   19-FEB-1998.
PD
XX
XX   15-AUG-1997; 97WO-US14436.
PF
XX
XX   16-AUG-1996; 96US-0024022.
PR
XX
XX   (SMIK ) SMITHKLINE BEECHAM CORP.
PA
XX
XX   Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI   Stodola RK;
XX
XX   WPI; 1998-159452/14.
DR
XX   P-PSDB; AAY85963.
XX
XX   Streptococcus pneumoniae proteins and related DNA - useful for
PT   screening compounds for antibacterial activity
XX
XX   Claim 4; Page 175; 640pp; English.
XX
XX   This invention describes novel isolated Streptococcus pneumoniae
CC   polynucleotides (see AAZ96173-296494) and their encoded proteins (see
CC   AAY85792-786182). The DNA, vectors and host cells described in the
CC   method of the invention are useful for the recombinant expression of the
CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   disease, or diagnosis of disease related to expression or activity of
CC   such a polypeptide. They can also be used to screen for compounds which
CC   interact with and inhibit or activate such a polypeptide. The
CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   for inducing an immunological response in a mammal. The antagonists are
CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   They are also useful to determine their role in pathogenesis of
CC   infection, dysfunction and disease.
XX
XX   Sequence 1729 BP; 546 A; 328 C; 336 G; 517 T; 2 other;
SQ
XX
XX   alignment_scores:
XX       Quality:      8.00      Length:      8
XX       Ratio:        1.000      Gaps:        0
XX   Percent Similarity: 100.000   Percent Identity: 100.000
XX
XX   alignment_block:
XX   US-09-697-089-2 x AAZ96314/rev ..
XX
XX   Align seg 1/1 to reverse of: AAZ96314 from: 1 to: 1729
XX
XX   688 Thrsrleuargleuglnlelyys 695
XX   ||||||||||||||||||||
XX   1645 ACATCTTGAGGTTGCAGATAAA 1622
XX
XX   seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH14793
XX   seq_documentation_block:
```

```
ID   AAH14793 standard: cDNA; 1729 BP.
XX
XX   AAH14793;
AC
XX
XX   26-JUN-2001 (first entry)
DT
XX
DE   Human cDNA sequence SEQ ID NO:12578.
XX
XX   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KM
XX
XX   Homo sapiens.
OS
XX
XX   EP1074617-A2.
PN
XX
XX   07-FEB-2001.
PD
XX
XX   28-JUL-2000; 2000EP-0116126.
PF
XX
XX   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
XX   (HELI-) HELIX RES INST.
PA
XX
XX   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX   WPI; 2001-318749/34.
DR
XX
XX   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
XX   Claim 8: SEQ ID 12578; 2537bp + CD ROW; English.
XX
XX   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences: AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
XX   SQ   Sequence 1729 BP; 563 A; 277 C; 354 G; 535 T; 0 other;
XX
XX   alignment_scores:
XX       Quality:      8.00      Length:      8
XX       Ratio:        1.000      Gaps:        0
XX   Percent Similarity: 100.000   Percent Identity: 100.000
XX
XX   alignment_block:
XX   US-09-697-089-2 x AAH14793/rev ..
XX
XX   Align seg 1/1 to reverse of: AAH14793 from: 1 to: 1729
```

895 SerLeuSerSerLeuLeuLysHis 902
 |||||
 1651 TCTTATCATGACCTTCTTAACAC 1628

seq_name: /SIDS2/gcdata/geneseq/geneseq/NM1999.DAT:AAZ24816

seq_documentation_block:

ID AAZ24816 standard; DNA; 1810 BP.

AAZ24816;

02-DEC-1999 (first entry)

Human secreted protein gene 6 clone HLGAM28.

Human; secreted protein; fusion protein; gene therapy; protein therapy;
 diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
 cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

MO9947540-A1.

23-SEP-1999.

18-MAR-1999; 99WO-US05804.

19-MAR-1998; 98US-0078563.
 19-MAR-1998; 98US-0078566.
 19-MAR-1998; 98US-0078573.
 19-MAR-1998; 98US-0078574.
 19-MAR-1998; 98US-0078576.
 19-MAR-1998; 98US-0078577.
 19-MAR-1998; 98US-0078578.
 19-MAR-1998; 98US-0078579.
 19-MAR-1998; 98US-0078581.
 01-APR-1998; 98US-0080312.
 01-APR-1998; 98US-0080313.
 01-APR-1998; 98US-0080314.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
 Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
 Olsen HS, Shi Y, Moore PA;

WPI: 1999-562050/47.
 P-PSDB; AAY41313.

New isolated human genes, useful for diagnosis and treatment of e.g.
 cancers, neurological disorders, immune diseases, inflammation or blood
 disorders

Claim 1; Page 302; 484pp; English.

This sequence represents a nucleic acid molecule which encodes a
 secreted human protein. The gene number, and the clone it is derived
 from, are detailed in the descriptor line. The gene can be used to
 generate fusion proteins by linking to the gene to a human immunoglobulin
 Fc portion (e.g. AAZ24802) for increasing the stability of the fused
 protein as compared to the human protein only.
 The invention relates to 95 novel genes and their fragments (nucleic
 acid sequences: AAZ24811-24907; amino acid sequences AAY41308-Y41404)
 which are useful for preventing, treating or ameliorating medical
 conditions e.g. by protein or gene therapy. Also, pathological
 conditions can be diagnosed by determining the amount of the new
 polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 95
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAZ24811 for described uses).

XX Sequence 1810 BP; 395 A; 479 C; 463 G; 473 T; 0 other;

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAZ24816

Align seg 1/1 to: AAZ24816 from: 1 to: 1810

703 SerLeuSerLeuValLeuSerThr 710
 |||||
 328 AGCCTCTCCCTGTCTCTAGCAGC 351

seq_name: /SIDS2/gcdata/geneseq/geneseq/NM1993.DAT:AAQ42858

seq_documentation_block:

ID AAQ42858 standard; DNA; 1839 BP.

AAQ42858;

28-SEP-1993 (first entry)

Serratia esterase.

Sr41; M1; mutation; allele; variation; hydrolysis; ss.

Serratia marcescens M1.

key Location/Qualifiers

FT mutation 9 /*lag= a

FT note= "T (M1) -> C (Sr41)."

EP544250-A.

02-JUN-1993.

25-NOV-1992; 92EP-0120078.

25-NOV-1991; 91JP-0355440.

13-APR-1992; 92JP-0137502.

(TANA) TANABE SEIYAKU CO.

Akatsuka H, Komatsubara S, Omori K, Shibatani T;

WPI: 1993-177043/22.

P-PSDB; AAR36774.

New genes encoding esterase from *Serratia* - are contained in
 recombinant vectors and transformed cells with high esterase
 productivity, useful as hydrolytic reagent

Claim 2; Page 10-11; 16pp; English.

When *S. marcescens* Sr41 was subjected to mutagenesis with
 N-methyl-N'-nitro-N-nitrosoguanidine and mutants selected on
 tributyrin medium, the mutant M1 was identified. M1 contains
 an allelic variant of C -> T at position 9. M1 has 2.5 times
 greater esterase prod. than Sr41. The esterase is useful for
 carrying out hydrolytic reactions.

Sequence 1839 BP; 386 A; 585 C; 536 G; 332 T; 0 other;

```
alignment_scores:
  Quality: 8.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Percent Identity: 100.000
alignment_block:
  US-09-697-089-2 x AAQ42858 ..
Align seg 1/1 to: AAQ42858 from: 1 to: 1839
450 ThrAlaGlyArgArgLeuSerSer 457
|||||
1094 ACAGCGGGCCGACGTTATCATCG 1117
```

THIS PAGE BLANK (USPTO)

OM of: US-09-697-089-2 to: GenEmbl:* out_format : pfs
Date: Mar 26, 2002 8:03 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=fastap+g2n.model -DEV=xlh
-O=/g9n2.1/USPO.spool/US09697089/runat_25032002.105211.12383/app-query.fasta.1.1097
-DB=genembi -Qfmt=fastap -SUFFIX=oligo.rge -GAPD=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPD=4.500 -OGAPEXT=0.050 -XGAPD=60.000 -XGAPEXT=60.000
-DELDP=6.000 -FGAPEXT=7.000 -YGAPD=60.000 -YGAPEXT=60.000
-DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=5 -MODE=LOCAL -OUTTEXT=pfs -NORM=ext
-HEA=SIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09697089 -EGCN_1.7624 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-697-089-2

Query length: 1024

Database: GenEmbl:*

Database sequences: 1472140

Database length: 341344837

Search time (sec): 1826.610000

WARN: XGAPD and YGAPD must be equal. Assuming YGAPD=XGAPD=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Score List:

Sequence	Strid Orig	ZScore	EScore	Len	Documentation
gb.pr:AY032589	+ 1024.00	19412.85	0.0	3133	AY032589 Homo sapiens caspase
gb.pr:AY035391	+ 923.00	14937.66	0.0	3219	AY035391 Homo sapiens ICE-1
gb.pr:AF376061	+ 822.00	15573.88	0.0	3381	AF376061 Homo sapiens caspase
gb.pr:AY027787	+ 721.00	13655.28	0.0	3396	AY027787 Homo sapiens CLAN
gb.pr:AC010968	+ 589.00	11118.93	0.0	160583	AC010968 Homo sapiens chrom
gb.pr:CN001533	- 559.00	10550.07	0.0	189909	AL121653 BAC sequence from
gb.pr:IR200517	+ 346.00	6537.07	0.0	1335	AL389934 Homo sapiens mRNA
gb.pr:AY027789	+ 271.00	5111.86	1.9e-276	1395	AY027788 Homo sapiens CLANB
gb.pr:CN001533	+ 154.00	2893.26	7.0e-153	768	AY027789 Homo sapiens CLANB
gb.pr:CN001533	- 97.00	1770.98	3.2e-90	165692	AL121658 BAC sequence from
gb.pr:AY027780	+ 89.00	1660.36	3.3e-84	578	AY027790 Homo sapiens CLAN
gb.pr:G55568	- 56.00	1033.69	2.7e-49	553	G55568 SHCC-100923 Human
gb.pr:AP00783	- 10.00	120.49	195.46	114233	AP00783 Homo sapiens genom
gb.pr:AC080039	+ 10.00	118.09	266.01	158375	AC080039 Homo sapiens chrom
gb.pr:AC080627	+ 10.00	117.64	281.78	168346	AC080627 Homo sapiens chrom
gb.pr:AC087661	+ 10.00	117.29	294.78	176593	AC087661 Homo sapiens chrom
gb.pr:AC011006	- 10.00	117.15	299.98	179901	AC011006 Homo sapiens chrom
gb.pr:AC048829	+ 9.00	147.40	6.19	222	AC048829 Rattus norvegicus
gb.pr:AC048829	+ 9.00	143.07	10.79	400	G24202 human STR W1-13519, sequ
gb.pr:HEP150	+ 9.00	140.11	15.79	599	L47474 Heterodontus francisci
gb.pr:AF030304	+ 9.00	135.64	28.02	11099	AF030304 Arabidopsis thaliana
gb.pr:AC051908	+ 9.00	135.64	28.02	11099	AC051908 Arabidopsis thaliana
gb.pr:AF181722	+ 9.00	133.96	34.75	1382	AF181722 Homo sapiens RUTAS
gb.pr:AF181722	+ 9.00	133.22	43.44	1751	X54359 P.sativum mRNA of CDNA
gb.pr:AF07036	- 9.00	131.23	49.32	2003	AF07036 Homo sapiens CDNA
gb.pr:AF181721	- 9.00	130.65	53.12	2167	AF181721 Homo sapiens RUTAS
gb.pr:AK027029	- 9.00	130.24	55.96	2290	AK027029 Homo sapiens CDNA
gb.pr:AF181720	+ 9.00	125.85	98.28	4161	J05443 Rat seminal vesicle
gb.pr:AF181720	+ 9.00	125.48	103.09	4377	AF181720 Homo sapiens RUTAS
gb.pr:AF181720	+ 9.00	119.43	224.05	9968	AF181720 Homo sapiens RUTAS
gb.pr:AE000560	- 9.00	119.15	232.16	10351	AE000560 Helicobacter pylori
gb.pr:AE000560	- 9.00	118.15	257.08	11533	AE000560 Pseudomonas aerugin
gb.pr:AC074110	- 9.00	114.12	442.46	20509	AC074110 Arabidopsis thaliana
gb.pr:AF078157	- 9.00	109.52	798.63	38358	AF078157 Caenorhabditis eleg
gb.pr:AF078157	- 9.00	109.37	813.35	39108	AF078157 Homo sapiens clef
gb.pr:AC079674	- 9.00	107.72	1.0e+03	48940	AC079674 Arabidopsis thaliana
gb.pr:AC083883	- 9.00	106.20	1.2e+03	60199	AC083883 Homo sapiens clone

gb.pr:AB007645	+ 9.00	104.81	1.5e+03	72698	AB007645 Arabidopsis thali
gb.pr:AC002122	+ 9.00	104.12	1.6e+03	79931	AC002122 Human unknown clo
gb.pr:FL1AD7	+ 9.00	103.97	1.6e+03	81513	AC021198 Sequence of BAC F
gb.pr:AF181721	+ 9.00	103.41	1.7e+03	88011	AL079344 Arabidopsis thali
gb.pr:AF181721	+ 9.00	102.42	2.0e+03	100625	AL079397 Human DNA sequen
gb.pr:AP001576	+ 9.00	102.10	2.1e+03	105199	AP001576 Homo sapiens gen
gb.pr:HS2747K22	+ 9.00	101.52	2.2e+03	113804	AL050333 Human DNA sequen
gb.pr:HS2747K22	+ 9.00	101.16	2.3e+03	119566	282195 Human DNA sequence
gb.pr:HS2747K22	+ 9.00	101.13	2.3e+03	120029	AL132672 Homo sapiens chr
gb.pr:HS2747K22	+ 9.00	101.01	2.4e+03	122014	AL139833 Human DNA sequen
gb.pr:AC003085	+ 9.00	100.84	2.4e+03	124836	AC003085 Human BAC clone
gb.pr:AC003085	+ 9.00	100.47	2.6e+03	134466	AC015608 Homo sapiens clo
gb.pr:AC010625	+ 9.00	100.07	2.7e+03	138638	AC010625 Homo sapiens chr
gb.pr:AC010625	+ 9.00	99.74	2.8e+03	145022	AC019177 Homo sapiens chr
gb.pr:AC027067	+ 9.00	99.67	2.8e+03	146403	AC027067 Homo sapiens chr
gb.pr:AL133550	+ 9.00	99.56	2.9e+03	148456	AL133550 Homo sapiens chr
gb.pr:AC091907	+ 9.00	99.49	2.9e+03	150037	AC091907 Homo sapiens chr
gb.pr:AL359713	+ 9.00	99.34	2.9e+03	152966	AL359713 Human DNA sequen
gb.pr:AL359713	+ 9.00	99.31	3.0e+03	153578	AC004074 Homo sapiens chr
gb.pr:AL359713	+ 9.00	99.26	3.0e+03	154715	AL359198 Homo sapiens chr
gb.pr:AC023214	+ 9.00	99.21	3.0e+03	155691	AC023214 Homo sapiens clo
gb.pr:AC022834	+ 9.00	99.21	3.0e+03	155833	AC022834 Homo sapiens chr
gb.pr:AP001444	+ 9.00	99.11	3.0e+03	157978	AP001444 Homo sapiens chr
gb.pr:AC055718	+ 9.00	99.07	3.0e+03	158746	AC055718 Homo sapiens chr
gb.pr:AC011838	+ 9.00	99.07	3.1e+03	158807	AC011838 Homo sapiens chr
gb.pr:AC004933	+ 9.00	99.06	3.1e+03	159082	AC004933 Homo sapiens chr
gb.pr:AC004933	+ 9.00	98.95	3.1e+03	161453	AC004933 Homo sapiens PAC
gb.pr:AC018601	+ 9.00	98.92	3.1e+03	162030	AC018601 Homo sapiens clo
gb.pr:AP002986	+ 9.00	98.92	3.1e+03	162066	AP002986 Homo sapiens chr
gb.pr:AC022488	+ 9.00	98.86	3.1e+03	162445	AC022488 Homo sapiens chr
gb.pr:AC011489	+ 9.00	98.83	3.1e+03	164034	AC011489 Homo sapiens chr
gb.pr:AC019132	+ 9.00	98.78	3.2e+03	165105	AC019132 Homo sapiens chr
gb.pr:AC025884	+ 9.00	98.76	3.2e+03	165341	AC025884 Homo sapiens chr
gb.pr:AC060768	+ 9.00	98.74	3.2e+03	165524	AC060768 Homo sapiens chr
gb.pr:AC021915	+ 9.00	98.74	3.2e+03	166039	AC021915 Homo sapiens chr
gb.pr:AC080132	+ 9.00	98.66	3.2e+03	167839	AC080132 Homo sapiens chr
gb.pr:AC023437	+ 9.00	98.57	3.3e+03	170022	AC023437 Homo sapiens chr
gb.pr:AC023437	+ 9.00	98.54	3.3e+03	170682	AC023437 Homo sapiens chr
gb.pr:AC025610	+ 9.00	98.53	3.3e+03	174715	AC025610 Homo sapiens chr
gb.pr:AC025610	+ 9.00	98.37	3.3e+03	174715	AC025610 Homo sapiens chr
gb.pr:AC087073	+ 9.00	98.36	3.3e+03	174821	AC087073 Homo sapiens chr
gb.pr:AC083880	+ 9.00	98.28	3.4e+03	176747	AC083880 Homo sapiens chr
gb.pr:AL603788	+ 9.00	98.24	3.4e+03	177668	AL603788 Homo sapiens chr
gb.pr:AC079640	+ 9.00	98.19	3.4e+03	178821	AC079640 Mus musculus clo
gb.pr:AC073583	+ 9.00	98.13	3.4e+03	180525	AC073583 Homo sapiens xp
gb.pr:AC026029	+ 9.00	98.09	3.5e+03	181309	AC026029 Homo sapiens chr
gb.pr:AC023487	+ 9.00	98.05	3.5e+03	182475	AC023487 Homo sapiens chr
gb.pr:AL021672	+ 9.00	98.04	3.5e+03	182736	AL021672 Homo sapiens chr
gb.pr:AL021672	+ 9.00	98.03	3.5e+03	182766	AL021672 Homo sapiens chr
gb.pr:AC084835	+ 9.00	98.03	3.5e+03	182964	AC084835 Homo sapiens chr
gb.pr:AC026614	+ 9.00	97.87	3.6e+03	187020	AC026614 Homo sapiens chr
gb.pr:AC084822	+ 9.00	97.79	3.6e+03	189016	AC084822 Homo sapiens chr
gb.pr:AC019227	+ 9.00	97.74	3.6e+03	190314	AC019227 Homo sapiens chr
gb.pr:AL050812	+ 9.00	97.66	3.7e+03	192208	AL050812 Homo sapiens chr
gb.pr:AC005296	+ 9.00	97.66	3.7e+03	192548	AC005296 Homo sapiens xp2
gb.pr:AC022544	+ 9.00	97.46	3.7e+03	197676	AC022544 Homo sapiens clo
gb.pr:AP000603	+ 9.00	97.37	3.8e+03	200000	AP000603 Homo sapiens gen
gb.pr:AL039183	+ 9.00	97.37	3.8e+03	200000	AL039183 Arabidopsis thal
gb.pr:AL039183	+ 9.00	97.32	3.8e+03	201512	AL039183 Human DNA sequen
gb.pr:AP001362	+ 9.00	97.27	3.8e+03	202777	AP001362 Homo sapiens chr
gb.pr:AC022338	+ 9.00	97.27	3.9e+03	202777	AC022338 Mus musculus clo
gb.pr:AC024116	+ 9.00	97.03	4.0e+03	209484	AC024116 Mus musculus chr

seq_name: gb.pr:AY032589

seq_documentation_block: 3133 bp mRNA
LOCUS AY032589 3133 bp mRNA 25-MAY-2001
DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete

ACCESSION AY032589
VERSION AY032589.1 GI:13899172

KEYWORDS
SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 3133)
 Gedges, B.J., Wang, L., Huang, W.-J., Lavelle, M., Manji, G.A., Brown, M., Jutman, M., Morganstern, J., Merriam, S., Glucksmann, A., DiStefano, P.S. and Bertin, J.
 Human CARD12 is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis
 Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001)

JOURNAL 11374873
 PUBMED 11374873
 2 (bases 1 to 3133)
 Bertin, J.
 Direct Submission
 Submitted (15-APR-2001) Neurobiology, Millennium Pharmaceuticals Inc., 640 Memorial Drive, Cambridge, MA 02139, USA

FEATURES
 Source
 1..3133
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 36..3110
 /note="CARD12: CED4/Apaf-1 family member"
 /codon_start=1
 /product="caspase recruitment domain protein 12"
 /protein_id="AAK38730.1"
 /db_xref="GI:1389173"
 /translation="MNFIDKNSRALIQMGMTVIKQITDDLFVNVNLNREEVNIICE KVEDDAARGIITHMLKKGSECNFLKSKEMNPLFODLNGOSLPHOTSGDDDLA ODKDIYTHPSLDFPYLGEDIDITFNKSTFTEPYLRKQHHRRVQDLNGLDIA LQSPCITBEGSGKSTLQRIAMGSGCKALKFKFVFLRLSRAGGLFETLCD QLDIPGTRKQTFMAMLLKLRQVLEFLDYGNEPFCPEIEALKENRFRKMWI VTTTECLRHIRQEGALAEVGMDEDAQALIRVLLKELAEGLLOIQSRCLRN MKTFLFVITCAIQMGSESEFHSHTQTLFHTFYDLQKHKHKGVAAPFIRSLDH CGDALCEGVFSHKFDELDQVSYNVEDLITGLCKYTAQRFKRYKFEFKRQDET ENYLEKQDNEALHETLIDRMNYLEQTLALMPMGCDYQGSISLKHLEEVQVLYGL KNNRLTDEIRILCAFRCNPLKFNQDLNLGNKYSSGGLMKFQFENLQDLVFEDF STKEFLPDPAVLRKLSOVLSKLTFLEARLVGMQFDDDLVITGAEFLVYA"

misc_feature
 36..299
 /note="CARD: Region: caspase recruitment domain"

misc_feature
 522..1406
 /note="NBS: Region: nucleotide-binding site"

misc_feature
 2001..3107
 /note="LRR: Region: leucine-rich repeats"

BASE COUNT 903 a 691 c 729 g 810 t

ORIGIN

alignment_scores:
 Quality: 1024.00 Length: 1024
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AY032589 ..

Align seg 1/1 to: AY032589 from: 1 to: 3133

1 MetAspPheIleIysAspAsnSerArgAlaIleuIIegIlnArgMetGlyme 17
 |||||
 36 ATGAATTTTATAAGACATACCGACCTTATTCAAAGATGGGAAT 85
 |||||
 17 tTtrValIleIysGlnIleThrAspAspLeuPheValTTrpAsnValIleuA 34
 |||||
 86 GACTGTATTAAAGCAATACACATGACCTATTGTATGAAATGTTCTGTA 135
 |||||
 34 snArgIuGluIuValAsnIleIleCysGluIuIysValGluGlnAspAla 50

|||||
 136 ATGCGCAAGACGTAACATCATTTGCTGCGAGAGGTGGCGAGATGCT 185
 |||||
 51 AlAArgGlyIleIleHtsMetIleLeuIysGlySerGluSerGlyAsp 67
 |||||
 186 GCTAGAGGGATCATTCACATGATTTTGAAAAGGGTTCCAGAGTCGTGTA 235
 |||||
 67 nLeuPheLeuIysSerLeuIysGluTTrpAsnTTrpProLeuPheGlnAsp 84
 |||||
 236 CCTCTTCTTAATCCCTTAGAGAGTGGAACTATCCTCATTTACAGACT 285
 |||||
 84 euAsnGlyIleSerLeuPheHisGlnThrSerGluIysAspAsp 100
 |||||
 286 TGAATGACAAAGTCTTTTCATCAGACATCAGAAAGAGCTTGACAGAT 335
 |||||
 101 LeuAlaGlnAspLeuIysAspLeuIysThrHisThrProSerPheLeuAsn 117
 |||||
 336 TTGGCTCAGATTAAAGACTTGTACCATACCCCATCTTTCTGTGAACCT 385
 |||||
 117 eTyProLeuGlyGluAspIleAspIleIlePheAsnLeuIysSerThr 134
 |||||
 386 TTTTCCCTTGATGAGATATTCACATATTTTAACTTCAAAAGACCT 435
 |||||
 436 TCACAGAACCTGCTGTGGAGAGACCAACACATCACCCTGAGAG 485
 |||||
 151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProGlyIle 167
 |||||
 486 CACCTACCTGATGAGCTCTGAGGCTCTTCAAGCCCTGACATCAT 535
 |||||
 167 eGluGlyGluSerGlyIysGlySerThrLeuLeuGlnArgIleAla 184
 |||||
 536 TGAAGGGAAATCGGCAAGGCAAGTCCATCGTCGACAGCGATTGCCA 585
 |||||
 184 eLeuThrGlySerGlyIysCysIysAlaLeuThrIysPheIysPheVal 200
 |||||
 586 TGCCTCGGGCTCCGGAAGTGCAGAGCTCTGACCAAGTCAATTCGTC 635
 |||||
 201 PhePheLeuArgLeuSerArgAlaGlnGlyIlePheGlnThrIle 217
 |||||
 636 TTTTCTCTCGCTCAGACAGGCGCCAGGGTGGACTTTTGAACCTCTG 685
 |||||
 217 sAspGlnLeuLeuAspIleProGlyThrIleArgIysGlnThrPheMet 234
 |||||
 686 TGATCAACCTCGATATACCTGCGCAATCAGAGAGACATTCATG 735
 |||||
 234 lAmetLeuLeuIysLeuArgGlnArgValLeuPheLeuLeuAspGlyTyr 250
 |||||
 736 CCATGCTGCGAGAGCTGCGGCAAGGGGCTCTTCTTGTGATGCTAC 785
 |||||
 251 AsnGluPheIysProGlnAsnGlyProGluIleGluAlaLeuIleIysG 267
 |||||
 786 AATGAATTCAGCCCGCAGACCTGCCAGAAATCGAACCTGTATTAAGCA 835
 |||||
 267 uAsnHisArgPheIysAsnMetValIleValThrThrThrThrGluCysL 284
 |||||
 836 AAMCACCCTGTCAGACATGTCATCGTCACACCATACACATGAGTGGC 885
 |||||
 284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
 |||||
 886 TGAGGCACATACCGCAGTTGTGGCTTGTACTGCTGAGGTGGGAGATATG 935
 |||||
 301 ThrGluAspSerAlaGlnAlaLeuIleArgIuValLeuIleIysGly 317
 |||||
 936 ACAAGAAACAGCCCGCAGGCTGTATCCGAGAGTGTGATCAAGAGGCT 985
 |||||
 317 uAlaGluGlyLeuLeuGlnIleGlnIysSerArgCysLeuArgAsn 334
 |||||
 986 TGCTGAAGGCTTGTCTCCAAATTCAGAAATCCAGGTGCTTGAAGATC 1035
 |||||
 334 euMetIysThrProLeuPheValValIleThrCysAlaIleGlnMetGly 350
 |||||

1036 TCATGAAGACCCCTCTTTGTGGTACATCACTTGCAATCCAGATGGCT 1085
 351 GluSerGluPheHisSerHisThrGlnThrLeuPheHisThrPheTy 367
 1086 GAAGTGTGTTCCACTCTCACACACAAACAGCGTTTCCATACCTCTTA 1135
 367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIas 384
 1136 TGATCTGTGTACAGAAACAAACAAACATAAAGGTGTGGCTGCAA 1185
 384 erAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
 1186 GTGACTTCATTCGGAGCCGTGGACCATGTGGAGACTAGCTCTGGAGGGT 1235
 401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValas 417
 1236 GTGTTCTCCCAAGATTGATTGAACTGCGAGATGTGTCACACGTGAA 1285
 417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysThrThrAlaGln 434
 1286 TGAGGATGTCCTCGTGAACACTGGGCTCCTCTGTAATATATACACTCAA 1335
 434 rGlyPheLysProLysTyrLysPhePheHisLysSerPheGlnGluTyrThr 450
 1336 GGTTCAGACCCAAAGTATAAATCTTTCACAAGTATTCACAGAGTACACA 1385
 451 AlaGlyArgAlaGluSerSerLeuLeuThrSerHisGluProGluGluVal 467
 1386 GCAGGACGACAGACTGACAGTTTATGACGTCATGACGACGACAGAGAGT 1435
 467 LThrLysGlyAsnGlyTyrLeuGlnLysMetValSerLieserAspIleT 484
 1436 GACCAAGGGAGATGTTACTTGCAAGAAATGTTCCATTCGGACATTA 1485
 484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
 1486 CATCACTTATAGCAGCCCTCGCGGTACACCTGTGGGTCACTGTGTGAA 1535
 501 AlaThrArgAlaValMetLysHisLeuAlaValTyrGlnHisGlyCys 517
 1536 GCCACGAGGCTGTATGAAACACCTCGCACAGATGATCAACACGGCTG 1585
 517 sLeuLeuGlyLeuSerLysLeuAlaLysArgProLeuTyrArgGlnLysSer 534
 1586 CCTTCTCGGACTTTCATCGCCAGAGGCTCTCTCGAGACAGAGATCTT 1635
 534 euGlnSerValLysAsnThrThrGlnGlnGluIleLeuLysAlaIleAsn 550
 1636 TGCAAAGTGTAAAAACACCACTGACGAAAGAAATTTGAAAGCCATAAAC 1685
 551 IleAsnSerPheValGluCysGlyIleHisLeuTyrGlnLysSerThrSe 567
 1686 ATCAATTCCTTTGTAGAGTGTGGCATCATTTATATCAAGAGATCAATC 1735
 567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSer 584
 1736 CAATACAGCCCTGAGCCCAAGAAATTTGAAGCTTTCTTCAAGGTAAACCT 1785
 584 euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGlu 600
 1786 TATATATCACTCAAGGAAATCCCGATTTACTTATTTGACTTCTTTGAA 1835
 601 HisLeuProAsnGlyAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
 1836 CATTTGCCCAATTTGTGCAAGTCTCTGGACTTCATTAACTGAGCTTTTA 1885
 617 rGlyGlyAlaMetAlaSerTyrGluLysAlaAlaGluAspThrGlyGly 634
 1886 TGGGGAGACTATGGCTTCATGGAAAAAGGCTGCAAGAACACAGAGTGAAA 1935
 634 leHisMetGluGluAlaProGluThrTyrIleProSerArgAlaValSer 650
 1936 TCCACATGGAGAGGCCCCAGAAACCTTACATTTCCAGCAGGCTGTATCT 1985

651 LeuPhePheAsnTyrLysGlnGluPheArgThrLeuGluValThrLeuAr 667
 1986 TTGTTCTTCAACGTGAGACAGAGATTCAGACTCTGGAGGTGCACACTCG 2035
 667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
 2036 GCATTTCAGCAAGTTGAATAGCAAGATATCATATCTGGGGAAAAATAT 2085
 684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 2086 TCAGCTTGCACCAAGCCTCAGGCTGCAGAAATTAAGATGTGCTGGTGTG 2135
 701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
 2136 CTGGAGCCCTCAGTTTGTGCTCCAGACCTGTGTAAGACATTTATCTCT 2185
 717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
 2186 CATGGTGAAGCCAGTCCCTCACCATAGAAATGAGAGGCACATCCAT 2235
 734 erValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
 2236 CTGTAAACAACCTGAAAACTTGAGTATTCATGACCTACAGAAATCAACGG 2285
 751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
 2286 CTGCCGGGTGTCTGACTGACAGCTTGGGTACTTGGAAGACCTTACAAA 2335
 767 sLeuIleMetAspAsnIleLysMetLysGluGluAspAlaIleLysLeu 784
 2336 GCTCATATATGATTAACATAAAGATGAATGAAGAAGATCTTAAACATAG 2385
 784 laGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
 2386 CTGAAGGCTGAAAAAACCCTGAAGAAGATGTTATTTCAATTTGACCCAC 2435
 801 LeuSerAspIleGlyGluGlyMetAspTyrIleValLysSerLeuSerSe 817
 2436 TTGTTCTGCAATTTGAGAGGAGGATGATTAATGATCAAGTCTGTGTACAG 2485
 817 rGluProCysAspLeuGlnGluIleGlnLeuValSerCysCysLeuSera 834
 2486 TGAACCTGTGACCTGTAAGAAATTCATTAAGTCCCTGCTGCTGTGTG 2535
 834 laAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
 2536 CAAATGCAGTAAAAATCTTACTCAGAAATCTTCAAAATTTGGTCAAAACCTG 2585
 851 SerIleLeuAspLeuSerGluAsnTyrLeuGlnLysAspGlyAsnGluAl 867
 2586 AGCATTTCTGATTTTATCAGAAAAATTAACCTGGAAGAAATGGAATGAAAGC 2635
 867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnLeuThrAlaL 884
 2636 TCTTCATGAACTGATCGACAGATGAACGTGCTGAAGACAGCTCACCGCAC 2685
 884 euMetLeuProTyrGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
 2686 TGATGCTGCCTCGGGCTGTGACGTGCAAGGACAGCTGAGAGCCCTGTGG 2735
 901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
 2736 AAACATTTGGAGAGAGTCCCAACACTGCTCAAGCTTGAGTTGAAAAACTG 2785
 917 parGluThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLys 934
 2786 GAGACTCAGACATACAGAGATTAGAAATTTAGGTGCAATTTTGTGAAGA 2835
 934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
 2836 ACCCTGTGAAAAACTCTCAGAGATTGAATTTGGGGGAAATCGTGTAGC 2885

```

951 SerAsGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuIsgIule 967
|||||
2886 AGTGAAGGATGGCTTGCCTTCATGGGTGATATTGAGAAATCTTAAGCAAT 2935
967 uValPhePheAspPheSerThrIlysgIupheLeuProAspProAlaLeu 984
|||||
2936 AGTGTGTTTGGACTTACTAGTAAAGAAATTTCTACCTGATCCAGCAATTAG 2985
984 aAlArgIysLeuSerGlnValIleuSerIlySerIlyThrPheLeuGlnGluAla 1000
|||||
2986 TCAGAAACTTAGCCAAAGTGTATCCAGATTACTTTCTGCAGAAAGCT 3035
1001 ArgLeuValAGIlyTrpGlnPheAspAspAspLeuSerValIleThrG1 1017
|||||
3036 AGCGTTGTTGGGTGGCAATTTGATGATGATGATCAGTGTATATACAG 3085
1017 yAlaPheIlyLeuValIThrAla 1024
|||||
3086 TGGCTTTAACTAGTACTGCT 3107

seq_name: gb_pr:AY035391

seq_documentation_block:
LOCUS AY035391 3219 bp mRNA PRI 24-JUL-2001
DEFINITION Homo sapiens ICE-protease activating factor mRNA, complete cds.
ACCESSION AY035391
VERSION AY035391.1 GI:14334214
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 3219)
Poyet,J.L., Srinivasula,S.M., Thnani,M., Razmaria,M.,
Fernandes-Alnemri,T. and Alnemri,E.S.
Identification of Ipadf, a human caspase-1-activating protein
related to Apaf-1
J. Biol. Chem. 276 (30), 28309-28313 (2001)
2 (bases 1 to 3219)
Poyet,J.L., Srinivasula,S.M., Fernandes-Alnemri,T. and
Alnemri,E.S.
Direct Submision
Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
USA

FEATURES
Source
1. 3219
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="p21-22"
145..3219
/note="IPAF; CED4/Apaf-1 family member; caspase-associated
recruitment domain containing protein"
/codon_start=1
/product="ICE-protease activating factor"
/protein_id="AAK59843.1"
/db_xref="GI:14334215"
/translation="MNFIKDNRALLIORMKNTVTKQITDDLFVNNVLNREVNIICCE
KVEDDARNGIITHMLKGSSECNLFKSLKWNVPLFODLNGOSLFQTSGLDDLA
QDLADYHTPSLFNPLGEBIDITFNLKSTFPPVPLRKQDHHRYEQLNLGDLA
LQSPCLIEGSGKSTLQRIAMLMGSKKALTKFVFLRLSRAQGLFETLQ
QLDIPGTRKOTPMAMLLKROVLFLDGNFKPQNCPEITALLKENHFRKNMVI
VTTEELRHROGALTAEGDMTESSAOLIREVLIKELAGLILQIOKSRCLRL
MKPLPYVITCALIOMGSESRSHQTTLPFTFYLLQCKKHKKKCYAASFISLKH
CGDLATGVFSHKDFELQDVSVNEVLLTTLGCLCTTAQREPKTKFKFSQETP
AGRLSSLTSHPEEVTNGKGYLQKNVSLSDITSTVSLRLRYCGSSVEATRAMKH
LAAVYQGLGLSLIAKRPRLROESLQSVKNTQDLIKAININSFVCCGTHLOEST
SKSLISOEFAPFGKSLIYSGNIPDYLPDFEHLNCAALDPKIDKIDFGAMASV
EKAEDTGGTHMEAPETIYPSRAVSLFNNKQFRPRLLEVTLRPSKLINQDRIYLG
IFSSATSLRLQIKRCAGVAGSLVLSLTKNITSLNWVASPLTEDEKHTITSVNLKT

```

```

LSTHDLQNRRLPGGLTDSLGNLKULTKLMDINKNNEEDAIKLEGLKMKLPHL
THLSIDGEMQYIYKSLSSPECDLEIQLVCCSAVAKILAOINLHVLTSLDLS
ENYLEKDNELHELIDRMNVLEQITLALMLPGCVQGSLSLKLHEEPQLYKLGL
KNWRLDREIRILICAFREKNPLKNKQINLAGNRRSSNGMLAFMGVFEENLKQLYPFD
STKEPLPALVRKLSYLSKTLTLOEARLVGMQFDDDDLSVITIGAFKLVTA"
BASE COUNT 934 a 696 c 748 g 841 t
ORIGIN

alignment_scores:
Quality: 923.00 Length: 1024
Ratio: 0.902 Gaps: 0
Percent Similarity: 99.902 Percent Identity: 99.902

alignment_block:
US-09-697-089-2 x AY035391
..
Align seg 1/1 to: AY035391 from: 1 to: 3219

1 MetAsnPheIleIlySAspAsnSerArGAlaLeuIleGlnArGmetGlyMe 17
|||||
145 ATGAATTTTCATTAAGGACATATACCCGACCCCTTATTCAAAGATGGGAAT 194
17 tThrValIleIlysgIlnIleThrAspAspLeuPheValTTPAsnValIleuA 34
|||||
195 GACTGTTATTAAGCAATCAAGATGACCTATTGTATGGAAGTTCTGTA 244
34 snArgGluGluValAsnIleIleCysGluGluIlyValGluGlnAspAla 50
|||||
245 ATCCGAAAGAGTAACATCATCTTGGCGAAGAGTGGAGCGAGTGGCT 294
51 AlaArgIlyIleIleHisMetIleLeuIlySlyGlySerGluSerCysAs 67
295 GCTAGAGGAGTATCATCATGATTTTGAAGAAAGGTTCAAGAGTCTGTAA 344
67 nLeuPheLeuIlySerLeuIlySglutrrPasnIlyrProLeuPheGlnAspL 84
|||||
345 CCTCTTCTTAAATCCCTTAAGAGTGAAGTAATCCCTCATTTTCAGAGCT 394
84 euAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
|||||
395 TGAATGACAAAGCTCTTTTCATCAGACATCAGAAAGACATCTTGACGAT 444
101 LeuAlaGlnAspLeuIlySAspLeuIlyrHisThrProSerPheLeuAsnPh 117
|||||
445 TTGGCTCAGAGATTAAAGGACTTGTACAAVACCCCATCTTTCTGAACTT 494
117 eTyrrProLeuGlyGluAspIleAspIleIlePheAsnLeuIlySerThrP 134
|||||
495 TTATCCCTTGGTGAAGATATTGACATTTATTTTAACTTGAAAGACACCT 544
134 heThrGluProValIleuTrpArgIlyAspGlnHisHisHisArgValGlu 150
|||||
545 TCACAGAACCTGCTCTGTGGAGAGAGACCAACCATCACCCTGGAGAG 594
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI1 167
|||||
595 CACCTGACCTGTAATGCGCTCTGACAGGCTCTTGACAGCCCTTCGACAT 644
167 eGluGlyGluSerGlyIlySlySerThrLeuLeuGlnArgIleAlaAla 184
|||||
645 TGAAGGGGAATCTGCCAAAGGCAAGTCCACTGCTGACGCGAATGGCA 694
184 eLeuTrpGlySerGlyIlyCysIlyAlaLeuThrIlySPhelysPheVal 200
|||||
695 TGGTCTGGGGCTCGGAAAGTGCAGAGCTTGACCAAGTTCAATTCGTC 744
201 PhePheLeuArgLeuSerArgAlaGlnGlyIlyLeuPheGluThrLeuCy 217
|||||
745 TTCTTCTCCTCGTGTACAGAGGCGCCAGGGGTGACCTTTTGAAACCTGCG 794
217 sAspGlnLeuLeuAspIleProGlyThrIleArgIlyGlnThrPheMetA 234
|||||

```

795 TGATCAACTCCTGGATATACCTGGCCACAATCAGAAAGCAGACATTCATCG 844
234 lAmMetLeuLeuLysLeuArgGlnArgValIleuPheLeuLeuAspLysTrp 250
|||||
845 CCATCTCTCTGAGCTGCGGAGAGGGTCTTCTCTTGTGATGGGTAC 894
251 AsnGluPheLysProGlnAsnGlyProGluIleGlnAlaLeuIleLysGlu 267
|||||
895 AATGAATTCAGAGCCCAAGACGCGCCAGAAATCGAAGCCCTGATTAAGA 944
267 uAsnHisArgPheLysAsnMetValIleValIleThrThrThrGluCysL 284
|||||
945 AAACCCAGCGCTTCAAGAACATGTCATCCACCTACCACTGATGCC 994
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
|||||
995 TGAGGCATATGCGGAGTTGGTGCCCTGACCTGAGAGTGGGGGAATAG 1044
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
|||||
1045 ACAGAAAGACAGCGCCAGCTCTCTCATCCGAGAGTGGTATCAAGAGACT 1094
317 uAlaGluLysLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
|||||
1095 TGCTGGAAGGCTTGTGCTCCAAATTCAGAAATCCAGGTGCTTGAAGATC 1144
334 eumMetLysThrProLeuPheValIleThrCysAlaIleGlnMetCysL 350
|||||
1145 TCATGMAAGACCCCTCTCTTGTGTCATCCTGTGCAATCCAGATGGGT 1194
351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
|||||
1195 GAAAGTGAGTCCACTCTCACACAAACAGCGTGTCCATACCTCTTA 1244
367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAlas 384
|||||
1245 TGATCTGTTGATACAGAAAAACAAACACAAACATTAAGTGTGGCTGCA 1294
384 eRAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
|||||
1295 GTGACTTCATTCGGAGCCTGGACACCTGTGGAGCCTTAGCTGTGAGGCT 1344
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValas 417
|||||
1345 GTGTTCCGCCCAAGTTTGATTCGAAGCTGCAAGATGTGTCACGCTGAA 1394
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyThrAlaGlnA 434
|||||
1395 TGAGGATGTCTCTGACAACTGGGCTCTCTGTAATATACAGCTCAAA 1444
434 rGpPheLysProLysTyLysPhePheHisLysSerPheGlnGluTyThr 450
|||||
1445 GGTTCAGACCCAAATATTAATCTTTTACACAGTCATTCAGAGATACACA 1494
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
|||||
1495 GCAGAGCAGAAAGACTCAGCAGTTTATGACGCTCATGACGACAGAGAGT 1544
467 lThrLysGlyAsnGlyTyLysLeuGlnLysMetValSerLieserAspLiet 484
|||||
1545 GACCAAGGGGAAATGTTACTTGAGAAAATGTTTCCATTTCCGACACTTA 1594
484 hrSerThrTySerSerLeuLeuArgTyThrCysGlySerSerValGlu 500
|||||
1595 CATCCACTTATAGCAGCGCTCTCGGTACACCTTGGGTCTATCTGTGAA 1644
501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyArgLnisGlyCys 517
|||||
1645 GCCACGAGGCTGTTATGAAAGACCTCGACAGAGTATCAACACGCGCTG 1694
517 sleuLeuGlyLeuSerLleAlaLysArgProLeuTPArgGlnGluSerL 534
|||||
1695 CCTTCTCGACTTTCATCGCAAGAGCGCTCTCTGAGACAGGAATCTT 1744
534 euGlnSerValLysAsnThrThrGluGlnGluIleLeuLysAlaIleAsn 550
|||||
1745 TGCAAGTGTGAAAAAACCCACTGAGCAGAAATTTGAAAGCCATTAAC 1794
551 IleAsnSerPheValGluCysGlyIleHisLeuTyArgGlnGluSerThrSe 567
|||||
1795 ATCAATTCCTTTGTAGAGTGTGGCATCCATTTATATCAAGAGATACATC 1844
567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSerL 584
|||||
1845 CAAATCACCCCTTGAGCCCAAGAAATTTGAAGCTTCTTTCAGAGTAAAGCT 1894
584 euTyrlIleAsnSerGlyAsnIleProAspTyrlLeuPheAspPheGlu 600
|||||
1895 TATATATCACTCAGGAGACATCCCGATTAATTAATTTGACTTCTTGAA 1944
601 HisLeuProAsnGlyAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
|||||
1945 CATTTGCCCAATTTGCAAGTGGCCCTGGACTTCATTTAACTGACCTTTTA 1994
617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyLysI 634
|||||
1995 TGGGGAGGCTATGGCTTCATGGGAAAAGGCTGCAGAAACACAGGTGGAA 2044
634 leHisMetGluGluAlaProGluThrTyrlIleProSerArgAlaValSer 650
|||||
2045 TCCACATGGAGAGAGCCCAAGAAACCTATACATTCACAGAGGGCTGTATCT 2094
651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValAlaThrLeuAr 667
|||||
2095 TTGTTCTTCAACTGAGAGCAGCATTCAGGACTCTGGAGGTACACATCCG 2144
667 gAspPheSerLysLeuAsnLysGlnAspLietThrTyLysGlyLysIleP 684
|||||
2145 GGATTCAGCAAGTTGATTAAGCAAGATATCAGATATCTGGGCAAAATAT 2194
684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
|||||
2195 TCAGCTCTGCCACNAAGCTCAGGCTGCAAAATTAAGATGGCTGGTGTG 2244
701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTySerLe 717
|||||
2245 GCTGGAAGCCTCAGTTTGTGCTCAGCACCGTGAAGAACTTTATTTCTCT 2294
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
|||||
2295 CATGGTGAAGCGAGTCCCTCACCATTAAGAAAGAGAGGACATTCACAT 2344
734 eValIThrAsnLeuLysThrLeuSerLleHisAspLeuGlnAsnGlnArg 750
|||||
2345 CTGTAAACAAACCTGAAACCTTGAGTATTCATGACCTACAGAAATCAACG 2394
751 LeuProGlyLysLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
|||||
2395 CTGGCGGGGTGTGACTGACAGCTTGGAACCTTGAAGAAACCTTAATAA 2444
767 sleuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeuA 784
|||||
2445 GCCTATATATGATTAACAATAAGATGAATGAAGAAAGATCATTAACCTAG 2494
784 laGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLleuThrHis 800
|||||
2495 CTGNAAGCCCTGAAAAACCTGAAGACAGTGTGTTATTTCAATTTGACCCAC 2544
801 LeuSerAspIleGlyGluGlyMetAspTyrlleValLysSerLeuSerSe 817
|||||
2545 TTGTTCTGACATTTGAGAGGGAATGGAATTAATAGTCAAGTCTGTGCAG 2594
817 rGluProCysAspLeuGluGluIleGlnLeuValSerCysCysLeuSera 834
|||||
2595 TGAACCTGTGACCTTGAAAGAAATTCATATAGTCTCTGCTGTGTCTG 2644

834 laasnaAlaVallyslleleuAlaGlnAsnLeuHisAsnLeuVallyslleu 850
|||||
2645 CAAATGCACTGAAATCCTAGCTCAGAAATCTTCACATTTGGCAAACTG 2694
851 SerIleLeuAspLeuSerGlnuAsnTyrlleuGlnLyAspGlyAsnGlnuAl 867
|||||
2695 AGCATTCTTGATTTATCAGAAAATTACCTGGAAAAAGATGGAATATCAAGC 2744
867 aleuHISgluLeuIleasparGMetAsnValleuGlnGlnLeuThrAla 884
|||||
2745 TCTTCATGAAGATGATGACAGATGAACGTCGAGAAACGATCCAGCCGAC 2794
884 eumeteuProtrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
|||||
2795 TGAATGCTGCCCTGGGGCTGTGACGTCCAAAGCGACGCTGAGACGCCCTGTTG 2844
901 LysHISleuGlnGlnValProGlnLeuVallyslleuGlnLyAsnTr 917
|||||
2845 AATCATTTGGAGAGGTCCACACACTGTCAGCTGGGCTGGAAGAACTG 2894
917 pArGleuThrAspThrGluIleArGlleuGlnLyAlaPhePheGlyLysA 934
|||||
2895 GAGACTCACAGATACAGAGATTAGAAFTTTAGTGATTTTGGAAAGC 2944
934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArGValSer 950
|||||
2945 ACCCTCTGAAAACTTCCAGCACTGAATTTGGCGGGAATCCTGTGAGC 2994
951 SerAspGlyTyrLeuAlaPheMetGlyValPheGlnAsnLeuLysGlnLe 967
|||||
2995 AGTGATGATGCTTCCCTTCATGAGGTGTATTGAGAAATCTTAAGCAATT 3044
967 uValPhePheAspPheSerThrLysGlnPheLeuProAspProAlaLeu 984
|||||
3045 AGCTTTTTCATTTAGACTTAAGAAATTTTCACTGATCCAGCACTTACG 3094
984 aAlArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGlnuAla 1000
|||||
3095 TCCGAAACCTTAGCCAAAGTGTATCCAAAGTTAACTTTTCGCAAGAACT 3144
1001 ArgLeuValGlyTyrGlnPheAspAspAspLeuSerValIleThrGln 1017
|||||
3145 AGCCTTGTTGGTGCGCAATTTGATGATGATGATCTCAGTCTTATTACAGS 3194
1017 yAlaPheLysLeuValThrAla 1024
|||||
3195 TCCTTTAACTAGTAAGTCT 3216
seq_name: gb_pr:AF376061
seq_documentation_block:
LOCUS AF376061 3581 bp mRNA PRI 15-MAY-2001
DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
ACCESSION AF376061
VERSION AF376061.1 GI:14040074
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 3581)
Gingras,M., Qiu,J. and Margolin,J.F.
TITLE Differential expression of the caspase recruitment domain protein 12 (CARD12) during monocytic differentiation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3581)
Gingras,M., Qiu,J. and Margolin,J.F.
AUTHORS Direct Submission
TITLE Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center,
JOURNAL Baylor College of Medicine, 6621 Fannin St. MC3-3320, Houston, TX
77030, USA
FEATURES
Location/Qualifiers

source 1. 3581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p21-22"
/cell_line="0937"
/cell_type="peripheral blood-derived CD14 mature monocyte"
490..3564
/note="CARD12"
/codon_start=1
/product="caspase recruitment domain protein 12"
/protein_id="AAK53443.1"
/db_xref="GI:14040075"
/translation="MNFIKDNRBALIORMGTATVTKQITDDLFWNNVLRREVNITICE
KVEDDARGIITHILKKGSECNLFSLKEMNPLPDODNGSLFPQTSGLDDDLA
ODLKDLYFTPSFLNFPYLDGIDIDIFNLKSTFPFVLMRDRDHHNRLEOLNGLD
LOSPLCITGEGSGKRSLLDRIAMLMGSKCKALKTKRPVPEFLSLSAOGSLPEPTLSD
QLDIPGTRKQTFMAMLIKROVNLFLDGYNEFKPQNCPEIALIKENHREKNMT
VTTEECLELRHROFGLTAELVAGDMTESSAQLIREVLIKELAEGLLIQIKSRCLNL
MKPLFVATICALIOMGESEPHSHQTLFTFYDILLQKNKHKKGVASDFTSLSDH
CGDALGVFSHKDFPELDVSVNEDALLTTGCLCYTAQREPKYKFPFKSFOEY
AGRLSSILTSHEPEEYTKNGYLOKKVSTISDITSTYSLRYTCGSSVEATRAVMK
LAAVYOHGCLIGLISIAKRPLMRQSLQSVKRTPEQTLKATININSFVCEGCHLYOEST
SKALSDPEFAFPOGKSLYINSNTIPYLPDFEFHLEPCASALDFIKLDYPGMAASH
EKAEEDVGIMHEAPEFYIPSRVAVSLFPMKQEFRLVEYLRDFSLNKODIRYLSK
IFSSATSLRLQIKRCAGVAGSLSLVSTCKNITYSLAMEASPLTIEDRHLTSTVNLKT
LSIHDLNORLPGLTDSLGNLKNLTKLIMONIKMNEBDAIKLAEGLNLKMLFHL
THLSDIGEMDYIVKSLSEPCDLEELIOLVSCCLSAANVRLAQNHLNLYLSTLDS
ENYLEKGNELHEHLIDRMNVLBDQTLMLPMWGDVGGSSLAEPGVCPEENIKOLYFPE
STRKFLDPALVRKLSOYLSKLRTPLQEARLYGWOFDDDLISYITGAKRLVYA"
502..741
misc-feature
BASE COUNT 1033 a 781 c 843 g 924 t
ORIGIN
alignment_scores:
Quality: 822.00 Length: 1024
Ratio: 0.804 Gaps: 0
Percent Similarity: 99.805 Percent Identity: 99.805
alignment_block:
US-09-697-089-2 x AF376061 ..
Align seg 1/1 to: AF376061 from: 1 to: 3581
1 MetAsnPhelIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLe 17
|||||
490 ATGAATTTTCATTAAGACATTAAGCCAGCCCTTATTCAAGAAATGGGAAT 539
17 tThrValIleLysGlnIleThrAspAspLeuPheValTyrPAsnValLeu 34
|||||
540 GACTGTTATTAACCAATCAACATGACCTATTGTATGAAATGCTTCTGA 589
34 snArgGluGlnValAsnIleIleCysCysGlnLysValGlnGlnAspAla 50
|||||
590 ATCGCGAAGAGTAACATCATTTGCTGCGAAGGTGAGCAGAGATGCT 639
51 AlArgGlyIleIleHisMetIleLeuLysGlnSerGlnSerCysAs 67
|||||
640 GCTAGAGGATCATTCATGATTTGAAAAGGTTCCAGAGTCTGTAA 689
67 nLeuPheLeuLysSerLeuLysGlnTyrPAsnTyrProLeuPheGlnAspL 84
|||||
690 CCTCTTTCTTAATCCCTTAAGAGAGTGAGACTTCTCTATTTCAGAGACT 739
84 euAsnGlnGlnSerLeuPheHisGlnThrSerGlnGlnLysAspAsp 100
|||||
740 TGAATGGACAAGACTTTTTCATCAGACATCAGAAGAGACTTGGAGAT 789
101 LeuAlaGlnAspLeuLysAspLeuTyrHisThrProSerPheLeuAsnPh 117
|||||

790 TTGGCTCAGGATTTAAAGACTGTGACCATCCCATCTTTCTGCACT 839
 117 eTyProLeuGlyIuAspIleAspIleIlePheAsnLeuLysSerThrP 134
 840 TTATCCCTGTGGTGAAGATATTGACATTATTTTAACTTGAAGACCACCT 889
 134 heThrGluProValLeuThrParGlyAspGlnHisHisIleArgValGlu 150
 890 TCACAGAACAGTCTGTGGAGGAAGACCAACACATCACCGCTGGAG 939
 151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
 940 CAGCTGACCTGAATGGCTCTCGAGGCTCTGAGAGCCCCCTGCATCAT 989
 167 eGluGlyGluSerGlyLysGlyLysSerThrLeuGlnIleArgIleAla 184
 990 TGAAGGGAAATGTGCNAAGGAAAGTCCATCTGTGCAGGCAATGGCA 1039
 184 eLeuThrGlySerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
 1040 TGCCTGGGGCTCCGGAAGTGCAGAGCTCTGACCAATTCATTCGCTC 1089
 201 PhePheLeuArgLeuSerArgAlaGlnGlyLysLeuPheGluThrLeu 217
 1090 TTCTTCCCTCCCTCAGCAGGCGCCAGGCTGACTTTTGAACCCCTCTG 1139
 217 sAspGlnLeuAsnSpleProGlyThrIleArgLysGlnThrPheMet 234
 1140 TGATCAACTCTGGATATACCTGGCACATCAGGAAGCAGACATTCATCG 1189
 234 lAmetLeuLysLeuArgGlnArgValLeuPheLeuLysPglYTr 250
 1190 CCATGCTGCTCAAGCTGGGCGAGGGTCTTTCTTCTTGATGGGCTAC 1239
 251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysG 267
 1240 AATGAATTCAGCCCAAGACTGCCAGAAATTCAGAGCCCTGATTAAGA 1289
 267 uAsnHisArgPheLysAsnMetValIleValThrThrThrGluCysL 284
 1290 AAACCCAGCGCTCAAGAACATGTGATGTGCACACTACCTAGTGC 1339
 284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
 1340 TGAGGCACATACGCGCAGTTGGTGCCTGCACTGCTGAGGTGGGAGATG 1389
 301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGlu 317
 1390 ACAGAAACAGCGCCAGGCTCTCATCCGAAAGTGTGATCAAGGAGCT 1439
 317 uAlaGlyGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsn 334
 1440 TGCAGGAGGCTGTGCTCCAAATTCAGAAATCCAGGTGCTTGAGGATC 1489
 334 euMetLysThrProLeuPheValIleThrCysAlaIleGlnMetGly 350
 1490 TCATGAGAGACCCCTCTTGTGTGTCATCTGTGCAATCCAGATGGCT 1539
 351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPhe 367
 1540 GAAAGTAGTCCACTCTCACACCAACAAACGCTGTTCATACCTCTCTA 1589
 367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIle 384
 1590 TGATCTGTGTACAGAAAAACAAACAAACATTAAGGTGTGGCTCAA 1639
 384 eArgPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGlyLys 400
 1440 GTGACTTCATTCGAGGCTGGACCATGTGAGACCTAGCTGTGAGAGGT 1689
 401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
 1690 GTGTTCTCCACAGATTGATTTCGAAGTGCAGAGATGTGTCCACCGTGA 1739

417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysThrAlaGln 434
 1740 TGAGATGCCCTGCTGACACTGGGCTCTCTGTGAATATATACACTCAA 1789
 434 rPheLysProLysTyLysPhePheHisLysSerPheGlnGlyTrThr 450
 1790 GGTTCAGGCCAAAGATATAAATTTTTCACAGTCACTTCCAGAGATACCA 1839
 451 AlaGlyArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
 1840 GCAGGACGAAAGCTCAGCAGTTTATTCACCTCATGAGCCAGAGAGGT 1889
 467 lThrLysGlyAsnGlyTyLeuGlnLysMetValSerIleSerSpleT 484
 1890 GACCAAGGGGATGTATCTTGCAAAAATGGTTTCCATTTCCGACATTA 1939
 484 hrSerThrTySerSerLeuLeuArgTyThrCysGlySerSerValGlu 500
 1940 CATCACTTATAGCAGCTGCTCGGTACACCTGTGGGTATCTGTGGAA 1989
 501 AlaThrArgAlaValMetLysHisLeuAlaValTyGlnHisGlyCys 517
 1990 GCCACAGGGCTGTATGAAACACCTCGCAGCACTGTATCAACAGGCTG 2039
 517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuThrArgGlnGluSer 534
 2040 CTTTCTCGACTTTCATCGCCAAAGAGCCCTCTGTGAGACAGGAATCTT 2089
 534 euGlnSerValLysAsnThrThrGlnGlnGluIleLeuLysAlaIleAsn 550
 2090 TGCAAAGTGTGMAAAACACCACTGACAGAAATTCGMAAACCATTAAC 2139
 551 lLeasnSerPheValGluCysGlyIleHisLeuTyGlnGluSerThrSe 567
 2140 ATCAATTCCTTTAGAGTGTGGCATCATTTATATCAAGAGATACATC 2189
 567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSer 584
 2190 CAATTCAGCCCTGAGCCACAGATTTGAACTTCTTTCAAGGTTAAAGCT 2239
 584 euTyrlLeasnSerGlyAsnIleProAspTyLeuPheAspPheGlu 600
 2240 TATATATCAACTGAGGACATCCCGATTAATTAATGACTTCTTGAA 2289
 601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPhe 617
 2290 CATTGGCCCAATTGTGCAAGTGCCTGACTTCATTAACTGACCTTTTA 2339
 617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyLys 634
 2340 TGGGGAGGCTATGGCTTCATGGGAAAAAGCTGCAGAAACACAGGTGAA 2389
 634 leHisMetGluGluAlaProGluThrTyrlleProSerArgAlaValSer 650
 2390 TCCACATGAGAAAGGCCCAAGAACTACATTCACACAGGCGGTATCT 2439
 651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeu 667
 2440 TTGTCTTCAACTGAGAGGAAATTCAGAGCTCTGAGGCTCACACCCG 2489
 667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyLeuGlyLysIleP 684
 2490 GGATTTACGAAAGTGAATTAAGCAAGATATACAGATATCTGGGAAAAAT 2539
 684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 2540 TCAGCTCTGCCACAAGCCTCAGGCTGCATAATAAGAGATGTGCTGGTGG 2589
 701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTySerIle 717
 2590 GTTGGAGGCTTCAGTTGTGTCCTCAGACACTGTAAAGAACTTTATTCCT 2639

```

717 uMeValGlAlAserProLeuThrILeGlUAspGluARHISILIEThrs 734
|||||
2640 CATGGTGGAGCCAGTCCCTCACCATGAGATGAGAGGACACACAT 2689
734 eValThraSnLeuLysThrLeuSerILeHISAspLeuGlnAsnGlna 750
|||||
2690 CTGTAAACAACCTGAAACCTTGAGATTATTCAGACCTACAGAAATCAACGG 2739
751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
|||||
2740 CTGGCGGGGGTCTGACTGACAGCTTGGGTACTGAAACCTTACAAA 2789
767 sLeuILeMetAspAsnILeLysMetAsnGluGluAspAlaILeLysLeuA 784
|||||
2790 GCTCATTAATGATTAACATTAAGATGATGAAAGAGTGTATTTCATTGACCCAC 2839
784 IaGlGLyLeuLysAsnLeuLysLysMetCysLeuPheHISLeuThrHS 800
|||||
2840 CTGAAGGCTGAAAACTGAAAGAAAGATGTGTATTTCATTGACCCAC 2889
801 LeuSerAspILeGlyGlyGlyMetAspTyrILeValLysSerLeuSerSe 817
|||||
2890 TTGTCTGACATTGGAAGAGGAATGATTACATAGTCAAGTCTCTGTCAAG 2939
817 rGluProCysAspLeuGluGluILeGlnLeuValSerCysCysLeuSerA 834
|||||
2940 TGAACCTGTGTGACCTTGAAGAAATTCAAATTAGTCTCTCTCTGTGTCTG 2989
834 IaAsnAlaValLysILeLeuAlaGlnAsnLeuHISAsnLeuValLysLeu 850
|||||
2990 CAAATGCAATGAAATCTTACGCTCAGAAATCTTCACAAATTTGGTCAAACTG 3039
851 SerILeLeuAspLeuSerGluAsnTyrLeuGluLysAspGlyAsnGluAl 867
|||||
3040 AACATTCTGATTATTATCAGAAAAATTTACTGAAAAAGATGAAATGAAACG 3089
867 aLeuHISGluLeuILeAspArgMetAsnValLeuGluGlnLeuThrAlaL 884
|||||
3090 TCTTCTAGACATGATCGACAGAGATGAACTGTGTAGAACAGCTCACCGCAC 3139
884 eUMetLeuProTrrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
|||||
3140 TGAATGCTGCGCTGGGGCTGTGACGTCAAGGACGCTGAGCACCTGTGTG 3189
901 LysHISLeuGluGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
|||||
3190 AAACATTTTGGAGAGGCTCCACAACTGCTCAAGCTTGGGTGAAAAACTG 3239
917 pArgLeuThrAspThrGluILeArgILeLeuGlyAlaPhePheGlyLysA 934
|||||
3240 GAGACTCACAGATACAGAGATTAGAAATTTAGGTGCAATTTTGGAAAAA 3289
934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
|||||
3290 ACCCTTGAAAAAATCTTCCAGCACTTGAATTTGGCGGGAAATCTGTGACG 3339
951 SerAspGlyTrrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
|||||
3340 AGTGAAGATGAGCTTGCCTTTCATGCGGTGATTTGACAAATCTTAAACAAT 3389
967 uValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeuV 984
|||||
3390 AGTGTATTTTGAATTAAGTAAAGAAATTTTACTGATCCAGCAATTTAG 3439
984 aLArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
|||||
3440 TCAGAAAACTTACCAAGTGTATTCAGATTACTTTTCTGCAAGAGCT 3489
1001 ArgLeuValGlyTrrpGlnPheAspAspAspLeuSerValILeThrG 1017
|||||
3490 AGGCTGTGTGGTGCAATTTGATGATGATGATCTCAGTCTTATTTACAG 3539
|||||
1017 'alaphelLysLeuValThrAla 1024

```

```

|||||
3540 TGCTTTAACTAGTACTGCT 3561
seq_name: gb_Pr:AY027787
seq_documentation_block:
LOCUS AY027787 3396 bp mRNA PRI 20-JUL-2001
DEFINITION Homo sapiens CLANA (CLAN1) mRNA, complete cds.
ACCESSION AY027787
VERSION AY027787.1 GI:14324112
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE Clana, a novel human ceo-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 3396)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
1..3396
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="p22-p21"
/tissue-type="lung"
1..3396
/gene="CLAN1"
277..3351
/gene="CLAN1"
/codon_start=1
/product="CLANA"
/protein_id="AAK14776.1"
/db_xref="GI:14324113"
/translation="MNFIKDNRALLIORMKMTYIKQITDLDLFPVNVNLRREVNICE
KVEDDAARGITIHMLKKGSESCNLEFLSKLKWNPPLPFDLNGSLPFOQTSGLDILA
QDLKDLVFPSPFLPGLDIDIFMLKSTFEPILMRDQHNHVRVOLTLGLDIA
LOSPLICGESGKSTLLDRIAMLSGSKALTKRFVFFPLRSPAOGLESTLGD
OLDIDPCTIKRPFAMLLKLRQVLELIDYNEFRQNPCELEALIKENHRKNMYI
VTTEELRHIRORGALTAEVGDWTESSQALIREVILKELAGLILIOKSCIRNL
MKPFLPVYITCAIDOMGSESRHQTTLFTFPDILLQKKHKHKGAAASDFIRSDH
RGDLALGVFSHKRDEFLQDVSVNEVLLTGLCKYTAQRKPKYKFEHKSFOEY
AGRLSILTSHEBEVTKNGVYLQKNVSIISDTSTYSSILARTCGSSVATRAVMKH
LAAVYOHGCLLGISIAKRPLMROESLQSVKNTPEOELIKAININSFVEGCIHLYOEST
SKSLISOEFEPFGKSLYINSGNIPYLFDEFPHLPNCASALDFIKLDYGGAMASM
EKAEDPTGIIHMEAPETYPISRAVSYFFNMKQEPFLLEVTLRDFSKLNODIRYLGK
IFSSATISLRQIKRCAGVAGSLSYLSTCKNTYSIMLEASPTLTEDBRNLTVYNLKT
LSIHDLQNRLLPGLDLSGLNKLNLKILMDNIKANEDEAIKLAEGNLKMKFLH
THLSIDIGENMEYIVKSLSEPCDELEIQLVSCCLSAVAIVILQNLNLKSLTDS
ENYLEKGENEADYHEHLIDRMVYLEQTLAMPWCGDVGSGSLKHLHEEYDQVYKGL
KNMLTLETEIRIILGAFGKNPLKNFOQLNLAGNVSSDGLAIFWGVENIKOLVFPDF
STKEFLDPALVRKLSQVLSKLTFLQEARLAVQWQFDDDDLTVITGARAKIVTA"
BASE COUNT 992 a 737 c 793 g 874 t
ORIGIN
alignment_scores:
Quality: 721.00 Length: 1024
Ratio: 0.706 Gaps: 0
Percent Similarity: 99.707 Percent Identity: 99.707
alignment_block:
US-09-697-089-2 x AY027787 ..
Align seg 1/1 to: AY027787 from: 1 to: 3396

```

```
1 MetAsnPhelIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyMe 17
|||||
277 ATGAAATTCCTAAGAGACAAATACAGCCACCCCTATTTCACAAATGGAGAT 326
|||||
17 tThrValIleLysGlnIleThrAspAspLeuPheValIlePheValLeuA 34
|||||
327 GACTGTATTAAGCAAAATCAACAGATGACATTTGATGACAAATGTTTCA 376
|||||
34 snArgLugIuValAsnIleIleCysCysGlyLysValIleGlnInspAla 50
|||||
377 ATGGGAAAGAAATAAACATCATTTGCTGCGAAGAGGTGGAGAGATCTT 426
|||||
51 AlaArgLylIleIleHisMetIleLeuLysLysGlySerGluSerCysAs 67
|||||
427 GCTAAGAGATCATTCACATGATTTTGAAAGAGGTTCAGAGTCTGTGAA 476
|||||
67 nLeuPheLeuLysSerLeuLysGluIleThrAspAspLeuPheGlnAspL 84
|||||
477 CCTCTTCTTAATCCCTTAAGAGGTGGAACTCTCTATTTCAGAGACT 526
|||||
84 euAsnGlyGlnSerLeuPheHisGlnIleThrSerGluLysAspLeuAsp 100
|||||
527 TGAATGGACAAAGCTTTTTCATCAGACATCAAGAGAGACTTGGACCAT 576
|||||
101 LeuAlaGlnAspLeuLysAspLeuIleThrProSerPheLeuAspH 117
|||||
577 TTGGCTCGAGATTTAAAGAGCTGTACACCCATCTTTCTGAACTG 626
|||||
117 eTyProLeuGlyLysAspIleAspIleIlePheAsnLeuLysSerThrP 134
|||||
627 TTATCCCTTGGAAGATATGACATTTATTTTAACTTGAAAGACACT 676
|||||
134 heThrGluProValLeuIleThrPArgLysAspGlnHisHisArgValG 150
|||||
677 TCACAGAACTATCTGTGGAGAGAGCAACACCACTGACCCGCTGAG 176
|||||
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIle 167
|||||
727 CAGCTGACCTGAATGGGCTCTCGAGAGCTTTCAGAACCCCTGCATAT 776
|||||
167 eGluGlyLysSerGlyLysSerThrLeuLeuGlnArgIleAlaIle 184
|||||
777 TGAAGGGAATCTGCAGAAAGCAAGTCCACTGCTGCGACCGCATGGCA 826
|||||
184 eLeuIlePheLysSerGlyLysCysLysAlaLeuThrLysPheLysPhe 200
|||||
827 TGGCTGGGGCTCCGAAAGATGCAAGGCTCTGACCAATTCACAAATTC 876
|||||
201 PhePheLeuArgLeuSerArgAlaGlnGlyLeuPheGlnIleThrL 217
|||||
877 TTCTTCCTCCCTCTCAGAGGCGCCAGGGTGACCTTTTGAAACCCTGT 926
|||||
217 snAspGlnLeuLeuAspIleProGlyIleArgLysGlnIleThrPheMe 234
|||||
927 TGATCAATCTCGATATACCTGCGACAAATCAAGAGACAACTTCATAG 976
|||||
234 ILeuIleLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTyr 250
|||||
977 CCACTCTCTGAAGCTGGGCGAGAGGGTCTTTCTTCCTTGATGGACAC 1026
|||||
251 AsnGluPheLysProGlnAsnCysProIleGlnAlaLeuLysGly 267
|||||
1027 AATGAATTCAGAGCCCAAGACTGCCAGAAATCGAAGCCCTGTAAAGA 1076
|||||
267 uAsnHisArgPheLysAsnMetValIleValIleThrThrThrGluCysL 284
|||||
1077 AAACCAAGCCCTCAGAAACATGGGTGATGTCACCACTACCAAGAGTCC 1126
|||||
284 euArgHisIleArgGlnPheGlyValAlaLeuThrAlaGluValGlyAsp 300
|||||
1127 TGAAGCAATACAGGAGATTTGGTGGCTGACATCTGAGAGGGGAGATG 1176
|||||

301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGlu 317
|||||
1177 ACAGAAACAGAGGCCAGGCTCTCTACCGAGAGAGTGTATCAAGAGAG 1226
|||||
317 nAlaGluGlyLeuLeuLeuIleGlnIlePheSerArgCysLeuArgAsnL 334
|||||
1227 TCGTGAAGCTTGTTGCTCCAAATTCAGAAATCCAGGCTTGAGAAATC 1276
|||||
334 euMetLysThrProLeuPheValIleIleThrCysAlaIleGlnMetGly 350
|||||
1277 TCATGAAGAGCCCTCTCTTTGTGCTCATCTGCACTTGCAATCCAGATG 1326
|||||
351 GluSerGluPheHisSerHisThrGlnIleThrThrLeuPheHisThrP 367
|||||
1327 GAAAGTGAATCTCACTCCACACAAACAGAGCGTTCACATACCTTCTA 1376
|||||
367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaHis 384
|||||
1377 TCATCTGTGATACAGAAAAACAACACAAACATTAAGGTGGCTGCAA 1426
|||||
384 eAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
|||||
1427 GTGACTCATTTGGAGCGCTGGACCCGCGAGACCTGACCTGAGAGGT 1476
|||||
401 ValIleSerHisLysPheAspPheGluLeuGlnAspValSerSerVal 417
|||||
1477 GTGTCTCCACAGATTTGATTTGATTCAGACCTGCAAGATGTGACAG 1526
|||||
417 nGluAspValLeuLeuThrThrGlyLeuLeuLysGlyThrPheAlaGln 434
|||||
1527 TGAAGATGTCTGCTGACACATGGGCTCTCTGTAAATATACACTCAA 1576
|||||
434 rGpHeLysProLysTyrLysPhePheHisLysSerPheGlnIleTyrThr 450
|||||
1577 GCTTCAGAGCAAGTAAATATTTTTCACAACTCATTTCCAGAGATTCACA 1626
|||||
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGlu 467
|||||
1627 GAGAGCAAGAGCTCAGAGATTTATACCTGTCAAGTGTGACAGAGAG 1676
|||||
467 IThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIle 484
|||||
1677 GACCAAGGGAATGTTACTTGCAGAAATGGTTTCATTTCCATTCGACAT 1726
|||||
484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValG 500
|||||
1727 CATCCACTATACAGAGCTGCTCGGATCACCTGTGAGATCTGATGAA 1776
|||||
501 AlaThrArgAlaValIleLysPheIleLeuAlaValIleThrGlnHisG 517
|||||
1777 GCCACAGAGGCTGTATGAAGGACCTGCGAGAGATGTATCAACAGGCT 1826
|||||
517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuIleThrPArgGlnSer 534
|||||
1827 CCTCTGAGACTTCCATCGCCAGAGAGGCTCTCTGAGACAGGAATCTT 1876
|||||
534 euGlnSerValLysAsnThrThrGlnGlnIleLeuLysAlaIleAsn 550
|||||
1877 TCCAAAGTGTGAAGAACCACTGAGCAAGAAATTCGAAACCCATTAAC 1926
|||||
551 rIleAsnSerPheValGlyCysGlyIleHisLeuTyrGlnIleSerPhe 567
|||||
1927 ATCAATTCCTTTGTAGAGGTGGCATTCATATATCAAGAGATACATC 1976
|||||
567 rLysSerAlaLeuSerGlnLysPheGlnAlaPhePheGlnGlyLysSer 584
|||||
1977 CAATTCAGCCCTGAGCCAGAAATTCGAACTTCTTTCAGAGGTAAAGCT 2026
|||||
584 euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGln 600
|||||
2027 TATATATCAACTCAGAGAGACATCCGCAATTCATATATTCATCTTGA 2076
|||||
601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPhe 617
|||||
```



```

|||||
2077 CATTTCGCCCAATTGTGCAAGTGCCTGACCTCATTTAACTGACCTTTA 2126
617 rglgylalamealasertrpbulysalaalaglualaspthrglyl 634
|||||
2127 TGGGGAGAGCTATGCTTCATGGGAAAAGCTGAGAAAGACACAGGTGGAA 2176
634 lehismetgluualaprogluThrlyleproserargalalaser 650
|||||
2177 TCCACATGGAAGAGGCCAGAAACCTACATTCGCCAGAGGCTGTATCT 2226
651 leuphephasntrpysglnupheargthrleugluvalthrleuar 667
|||||
2227 TTGTTCTCACTGGAAGAGGAATTCAGACTCTGAGGTCACTCCG 2276
667 gaspserlysleuasnlsglnaspllethrlyleuglylslel 684
|||||
2277 GGATTTCACAGCAAGTTGAATAGCAATATCAGATATCTGGGAAATAT 2326
684 heserSerAlathrSerleuArgleuGlnlelysarGysalaglual 700
|||||
2327 TCAGCTCTGCACACAGCTCAGGCTCAATAAAGAGATGCTGCTGTG 2376
701 AlaGlySerleuSerleuValleuSerThrCysGysAsnlelyrserie 717
|||||
2377 GCTGGAGAGCTCAGTTGTGCTCAGACACCTGTAAGAACATTTATCTCT 2426
717 uMetValGluAlaSerProleuThrIleGluAspGluArgHisIleHs 734
|||||
2427 CAGTGGGAAGACCCAGTCCCTCACCATAGAAATGAGAGGACATCAT 2476
734 erValIthrAsnleuLythrleuSerIleHisAspleuGlnasnglna 750
|||||
2477 CTGTAAACAACCTGAAACCTGTAGATTATCATACCTACGAAATCAAC 2526
751 leuProGlyGlyleuThrAspSerleuGlyAsnleuLysasnlleuHrly 767
2527 CTGCCGGGTGTGCTGACTGACACTGGGTAACCTTGAAGACCTTACAA 2376
767 sleuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysleu 784
|||||
2577 GCTCATTAATGATACATTAAGATGAATGAAGAATGCTATAAACTAG 2626
784 lagluGlyleuLysAsnleuLysLysMetCysleuPheHisleuThrHis 800
2627 CTGAAGGCTGAAACCTGAAAGACATGCTTATTATTATTGACCCAC 2676
801 leuSerAspIleGlyGlyMetAspIyrIleValLysSerleuSerse 817
|||||
2677 TTGTTGACATTGAGAGGGAATGATTACATATGTCAGTCTGTCAG 2726
2727 TGAACCTGTGACCTTGAAGAAATTCATATAGTCTGCTGCTGCTG 2776
834 laAsnAlaValLysIleleuAlaGlnAsnleuHisAsnleuValLysleu 850
|||||
2777 CAATTCACAGGAAATCTAGCTCAGAAATCTTCACAAATTTGGTCAAA 2826
851 SerIleleuAspLeuSerGluAsnIyrleuGluLysAspGlyAsnGlu 867
2827 AGCATTTCTGATTATTCAGAAATTCCTGCGAAGAAAGATGGAATGAG 2876
867 aleuHisGluLeuIleAspArgMetAsnValleuGluGlnleuThrAla 884
2877 TCTTCATGAAGCTGATGACAGATGAAGCGTGAAGAACAGCTACCGC 2926
884 eumetleuProtrpGlyCysAspValGlnGlySerleuSerleu 900
2927 TGTATGTCGCTGCGGCTGTGACGTGCAAGGCGCTGAGACACCTGCT 2976
901 LysHisleuGluGluValProGlnleuValLysleuGlyleuLysAsn 917
|||||

```

```

2977 AAACATTTGGAGAGGCTCCACAACTGTCAGCTTGGGTGAAAAACTG 3026
917 parGleuThrAspThrGluIleArgIleleuGlyAlaPhePheGlyLys 934
|||||
3027 GAGACTCACAGATACAGACATTAAGATTTAGTGCATTTTGGAAAGA 3076
934 snProleuLysAsnPheGlnGlnleuAsnleuAlaGlyAsnArgValSer 950
|||||
3077 ACCCTCGAAAAAATCTCCAGCAGTTGAATTTGCCGGAATCGTGGAC 3126
951 SerAspGlyThrleuAlaPheMetGlyValPheGluAsnleuLysGln 967
|||||
3127 AGTATGATGATGCTGCTTCATGGGTGATTGAGAAATCTTAAGCAAT 3176
967 uValPhePheAspPheSerThrLysGlnPheleuProAspProAlaLeu 984
|||||
3177 AGGTGTTTTGACTTATGACTTAAGAAATTTCTACTGATCCAGCATTG 3226
984 aArgLysleuSerGlnValleuSerLysleuThrPheleuGlnGluAla 1000
|||||
3227 TCAGAAACTTAGCCCAAGTGTATCCAAAGTAACTTTCTGCAAGAGCT 3276
1001 ArgleuValGlyThrGlnPheAspAspAspAspAspLeuSerValIleThr 1017
|||||
3277 AGCTTGTTGGGTGGCAATTTGATGATGATGATGATGATGATGATGAT 3326
1017 yAlaPheLysleuValThrAla 1024
|||||
3327 TGCTTTAAACTAGTAAGTCT 3348

```

seq_name: gb_htg:AC010968

seq_documentation_block:

LOCUS AC010968 160583 bp DNA HTG 18-AUG-2000
 DEFINITION Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE,
 11 unordered pieces.

ACCESSION AC010968.5 GI:9845170

VERSION AC010968.5 HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS 1 (bases 1 to 160583)

TITLE Waterston, R.H.

JOURNAL The sequence of Homo sapiens clone

REFERENCE Unpublished

AUTHORS 2 (bases 1 to 160583)

TITLE Waterston, R.H.

JOURNAL Direct Submission

REFERENCE Submitted (28-SEP-1999) Genome Sequencing Center, Washington

COMMENT MO 63108, USA

On Aug 18, 2000 this sequence version replaced gi:8439959.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0093002
 ----- Summary Statistics -----
 Sequencing vector: M13: 59k
 Sequencing vector: plasmid: 41k
 Chemistry: dye-primer ET: 47% of reads
 Chemistry: dye-terminator Big Dye: 53% of reads
 Assembly program: Phrap: version 0.990319
 Consensus quality: bases at least Q40
 Consensus quality: bases at least Q30
 Consensus quality: bases at least Q20
 Insert size: 14700; agarose-fp
 Insert size: 159583; sum-of-contigs
 Quality coverage: 6.64 in Q20 bases; agarose-fp

Quality coverage: 6.38 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1080: contig of 1080 bp in length
* 1081 1180: gap of unknown length
* 1181 2476: contig of 1296 bp in length
* 2477 2576: gap of unknown length
* 2577 5035: contig of 2459 bp in length
* 5036 5135: gap of unknown length
* 5136 9606: contig of 4471 bp in length
* 9607 17078: contig of 7372 bp in length
* 17079 17178: gap of unknown length
* 17179 27158: contig of 9980 bp in length
* 27159 27258: gap of unknown length
* 27259 45137: contig of 17879 bp in length
* 45138 45237: gap of unknown length
* 45238 65522: contig of 20285 bp in length
* 65523 91489: gap of unknown length
* 91490 91598: gap of unknown length
* 91599 116835: contig of 25237 bp in length
* 116836 160583: contig of 43648 bp in length.
* 116936 Location/Qualifiers

FEATURES
source 1. 160583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-9302"

misc_feature 1. 1080
/note="assembly_name:Contig9"
misc_feature 1181. 2476
/note="assembly_name:Contig12"
misc_feature 2577. 5035
/note="assembly_name:Contig13"
misc_feature 5136. 9606
/note="assembly_name:Contig14"
misc_feature 9707. 17078
/note="assembly_name:Contig15"
misc_feature 17179. 27158
/note="assembly_name:Contig16"
misc_feature 27259. 45137
/note="assembly_name:Contig17"
misc_feature 45238. 65522
/note="assembly_name:Contig18"
misc_feature 9607. 17078
/note="assembly_name:Contig19"
misc_feature 91599. 116835
/note="assembly_name:Contig20"
misc_feature 116936. 160583
/note="assembly_name:Contig21"
BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
ORIGIN

alignment_scores:
Quality: 589.00 Length: 589
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AC010968 ..

Align seg 1/1 to: AC010968 from: 1 to: 160583

89 LeupheHisGlnThrSerGluGlyAspLeuAspAspLeuAlaGlnAspLeu 105
|||||
49190 CTTTTCATCATGACATCCAGAGGAGACTTGGACGATTTGGCTCAGGATTT 49239
105 uLysAspLeuThrHisThrProSerPheLeuAspPheThrProLeuGly 122
|||||
49240 AAAGACTTGTATACCATACCCATCTTTCTGAACCTTTATCCCTGGTGG 49289
122 LuAspLeuAspLelIlePheAsnLeuLysSerThrPheThrGluProVal 138
|||||
49290 AAGATATTGACATATATTTTAACTTGAAGAACCTTACAGAACCTTCTC 49339
139 LeuThrParGlyAspGlnHisHisHisArgValGluGlnLeuThrLeuAs 155
|||||
49340 CTGTGGAGAGAGACCAACACCATCACCGCTGGAGAGCTGACCTGAA 49389
155 nGlyLeuLeuGlnAlaLeuGlnSerProCysLelIleGluGlyGluSerG 172
|||||
49390 TGGCTCTCTGACGCTCTTCAGACCCCTGCATCATTTGAAGGGAATCTG 49439
172 LysGlyLysSerThrLeuLeuGlnArgLleAlaMetLeuThrGlySer 188
|||||
49440 GCAAGGCAATCCACTCTGCTGACGCGCATTTGCCATGCTTGGGGCTCC 49489
189 GlyCysLysAlaLeuThrLysPheLysPheValPhePheLeuArgLe 205
|||||
49490 GGAAGTCAAGGCTCTGACCAAGTCAATTCGTCCTTCCTCCCTGCT 49539
205 uSerArgAlaGlnGlyGlyLeuPheGlnThrLeuCysAspLlnLeuAla 222
|||||
49540 CAGAGGGCCCAAGGTGACTTTTGAACCTCTGTATCAACTCTCTGG 49589
222 sPleProGlyThrLleArgLysGlnThrPheMetAlaMetLeuLeuLys 238
|||||
49590 ATATACCTGGCACAATCGAAGCAGACATTCATGGCCATGCTCTGTAG 49639
239 LeuArgGlnArgValLeuPheLeuLeuAspGlyTyrAsnGluPheLysPR 255
|||||
49640 CTCGGCAGAGGCTTCTTCTCTGATGCTACATCAATCAAGCC 49689
255 oGlnAsnCysProGlnIleGlnAlaLeuLleLysGlnAsnHisArgPheL 272
|||||
49690 CCAGAACTGCCCCAGAAATCGAAGCCCTGATTAAGGAAACCCAGCTTCA 49739
272 yAsnMetValIleValThrThrThrThrGluCysLeuArgHisIleArg 288
|||||
49740 AGAACATGCTCATGCTCACCCATCACCTGATGCTGAGGACATACCG 49789
289 GlnPheGlyAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAl 305
|||||
49790 CAGTTTGGTGGCTGACCTGACCTGAGGTGGGATATGACAGAACAGACGC 49839
305 aGlnAlaLeuLeuArgGluValLeuLleLysGlnLeuAlaGluLysLeu 322
|||||
49840 CCAGGCTCTCATCCGAGAAAGTGTATCAAGGACCTTGCTGAAGGCTTGT 49889
322 euleGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 338
|||||
49890 TGCCTCCAAATTCAGAAATCCAGGTGCTTGAAGATTCATCAAAACCCCT 49939
339 LeuPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 355
|||||
49940 CTCTTTGTGTCATCACTTGTGCAATCCAGATGAGGTGAAGTGAAGTCA 49989
355 sSerHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleG 372
|||||
49990 CTCTCACACAAACAAAGCGCTGTCCATACCTTTATGATCTGTGTGATAC 50039
372 InLysAsnLysHisLysGlyValAlaAlaSerAspPheIleArg 388

```

50040 AGAAAAACAACAAACATATAAAGGTGTGCTCAAGTACTGCTTCAATTCGG 50089
389 SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisIsty 405
50090 AGCCTGACACACTGTGAGACCTAGCTGTGAGGTGTGTCTCCACAA 50139
405 sPheAspPheGluLeuGlnAspValSerSerValAsnGluAspValLeuL 422
50140 GTTGAATTCGAACTCGAGAGTGTCCAGCCGTGAATGAGAGTGTCTCG 50189
422 eutHrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPheLysProLys 438
50190 TGCACTGGGGCTCCTCTGTAATATACAGCTCAAGGTTCAAGCCCAAG 50239
439 TyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGlyArgArgle 455
50240 TATAAATCTTTCACAAATGATTCATTCAGAGTACACAGCAGCAGAGAGCT 50289
455 uSerSerLeuLeuThrSerHisGluProGluGluValThrLysGlyAsnG 472
50290 CAGCAGTTATTACGCTCTCATAGACCCAGAGGAGGTGACCAAGGGAATG 50339
472 LyrLeuGlnLysMetValSerLysSerAspLethrSerThrTyrSer 488
50340 GTTACTTGCAGAAATGTTTCATTCGAGACTTACATCCACTTATAGC 50389
489 SerLeuLeuArgTyrThrCysGlySerSerValGluAlaThrArgAlaya 505
50390 ACCCTGCTCCGGGTACACCTGTGGTCATCTGTGGAAGCACCAGGGCTGTG 50439
505 MetLysHisLeuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLys 522
50440 TATGAAGCACCTCGCAGAGTATATACACGGCTCCCTTCGTGGACTTT 50489
522 erLleAlaLysArgProLeuThrParGlnGluSerLeuGlnSerValLys 538
50490 CCATGCCCAAGAGGCTCTCTGTGAGACAGGAATCTTGGCAATGTGGA 50539
539 AsnThrThrGluGlnGluLeuLysAlaLleAsnLleAsnSerPheVa 555.
50540 AACACACACGAGCAGCAAAATTCGAAAGCCATMAACATCAATCCCTTGTG 50589
555 LgluCysGlyLleHisLeuTyrGlnLysSerThrSerLysSerAlaLeuS 572
50590 AAGAGTGTGCATTCATTATATCAGAGAGTACATCCAATTCAGCCCTGA 50639
572 erGlnGluPheGluAlaPhePheGlnGlyLysSerLeuTyrLleAsnSer 588
50640 GCCAAGATTTGAGCTTCTTTCAGAGTAAAAAGCTTATATATACACTCA 50689
589 GlyAsnLleProAspTyrLeuPheAspPhePheGlnHisLeuProAsnCy 605
50690 GGGAAATCCCGGATTAATTTGACTTCTTGAACATTTGCCCAATTG 50739
605 sAlaSerAlaLeuAspPheLleLysLeuAspPheTyrGlyAlaMetA 622
50740 TGCAGTGTGCTGTGAGCTTCATTAAGCTTATATGAGGGGAGCTATGG 50789
622 LAserTProLysAlaAlaGluAspThrGlyGlyLleHisMetGluGlu 638
50790 CTTTCATGGGAAAAAGGCTGCAGAAACACAGGTGAGTCCACATGGAAG 50839
639 AlaProGluThrTyrLleProSerArgAlaValSerLeuPhePheAsnTr 655
50840 GCCCCAGAAACCTACATCCACAGCAGGCTGTATCTTTCTTCAACTG 50889
655 pLysGlnGluPheArgThrLeuGlnValThrLeuAspPheSerLysL 672
50890 GAAGCAGGAATTCAGACTCTGAGGTACACTCCGGATTTTCAGCAAGT 50939
672 euAsnLysGlnAspIle 677
|||||

```

```

50940 TGAATAGCAAGATATATC 50956
seq_name: gb_pr:CNS01DS3
seq documentation block:
LOCUS CNS01DS3 138909 bp DNA PRI 02-MAR-2000
DEFINITION BAC sequence from the SP64 candidate region at 2p21-2p22 BAC 164M19
of C17H_978_SKB library from chromosome 2 of Homo sapiens (Human),
complete sequence.
ACCESSION AL121653
VERSION AL121653.2 GI:7159616
KEYWORDS HTG: HTGS, DRAFT; Sep4 genomic DNA interval.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 138909)
AUTHORS Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D.,
Artiguenave, F., Davoine, C., S., Ciraud, C., Durr, A., Wincker, P.,
Brotier, P., Cactolico, L., Barbe, V., Burgunder, J. M.,
Prud'Homme, J. F., Brice, A., Fontaine, B., Hellig, R. and
Weissenbach, J.
TITLE Spastin, a novel AAA protein, is altered in the most frequent form
of autosomal dominant spastic paraplegia
JOURNAL Nat. Genet. (1999) In press
REFERENCE 2 (bases 1 to 138909)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (29-FEB-2000) to the EMBL/GenBank/DBJ databases
COMMENT On Mar 6, 2000 this sequence version replaced gi:6002386.
FEATURES
Source location/Qualifiers
1..138909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone_lib="C17H_978_SKB"
/clone="164M19"
BASE COUNT 39243 a 28424 c 29121 g 42121 t
ORIGIN
alignment_scores:
Quality: 589.00 Length: 589
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment block:
US-09-697-089-2 x CNS01DS3/rev ..
Align seg 1/1 to reverse of: CNS01DS3 from: 1 to: 138909
89 LeupheHisGlnThrserGluGlyAspLeuAspAspLeuAlaGlnAspIle 105
|||||
75307 CTTTTCATCAGACATCAGACAGAGAGCTTGAGAGATTTGGCTCAGGATTT 75258
105 uLysAspLeuTyrHisThrProSerPheLeuAsnPhetyrProLeuGlyG 122
|||||
75257 AAAGGACTTGACATATACCCCATCTTTCTGAACTTTATGCCCTGTGGTG 75208
122 LuAspLleAspLleLlePheAsnLeuLysSerThrPheThrGluProVal 138
|||||
75207 AGATATTACATTAATTTTAACCTGAAAAGCAGCTTCACAGAACCTGTG 75158
139 LeuThrParGlyLysAspGlnHisHisArgValGluGlnLeuThrLeuAs 155
|||||
75157 CTGTGAGAGAGAGACACACACATCCCGCTGGAGACACTGACCTGAA 75108
155 nGlyLeuLeuGlnAlaLeuGlnSerProCysLleLleGluGlyGluSerG 172
|||||
75107 TGGCTCTGTGAGGCTCTTACAGAGCCCTGCATCATTTGAAGGGGAATCTG 75058
172 LylsGlyLysSerThrLeuLeuGlnArgIleAlaMetLeuThrPglYser 188
|||||

```

75057 GCAAAGCAAGTCCACTCTGCTGCAGCGAAATGGCATGCTCTGGGGCTCC 75008
 189 G1ylyscyslysa1aleuthrlysphe1ysphe1val1phe1he1leu1arg1le 205
 75007 GGAAAGTGCAGAGGCTGTGACCAAGTCAAAATCTCTTCTTCCGCTCT 74958
 205 user1arg1al1agl1ng1ly1leu1phe1he1lu1thr1leu1cys1asp1leu1leu1a 222
 74957 CACAGGGGCCAGGGTGGACTTTTGAACCTCTGTGATCAATCCTCTGG 74908
 222 spl1eep1ogl1ythr1le1arg1lys1gl1n1thr1phe1met1al1met1leu1lys 238
 74907 ATATACCTGGCACATCAGAGACAGACATTCATGGCCATGCTCTCTGAAG 74858
 239 leu1arg1lu1n1arg1val1leu1phe1he1leu1asp1ly1tr1asn1gl1up1he1ly1spr 255
 74857 CTGGGGGAGAGGGTCTTCTTCTTCTTGTGATGGCTACAAATGAAATCAAGCC 74808
 255 og1n1asn1cys1pro1gl1u1le1gl1u1al1leu1le1lys1gl1u1asn1h1s1r1p1he1 272
 74807 CCAGAACTGCCCAAGAAATCCAGAGCCCTGATTAAGAAACCACCGCTTCA 74758
 272 ysa1sn1met1val1le1val1thr1thr1thr1gl1uc1ys1leu1arg1h1s1le1arg 288
 74757 AGAACAATGGTCATCGTCACCACTACCTAGTCCCTGAGGACATACGG 74708
 289 gl1p1he1gl1y1al1eup1h1r1al1agl1u1val1gl1y1asp1met1h1r1gl1u1n1s1p1ser1al 305
 74707 CAGTTTGGTCCCTGACTGCTGAGGGGGGATATGACAGAAAGCACAGCC 74658
 305 ag1n1al1al1eul1e1arg1lu1val1leu1le1lys1gl1u1leu1agl1u1gl1y1leu1 322
 74657 CCAGGCTCTCATCCAGAAATCCAGAGCTGATCAAGAGCTTCTGGAAGGCTTGT 74608
 322 eu1leu1gl1n1le1gl1n1lys1ser1arg1cys1leu1arg1asn1leu1met1ly1s1thr1pro 338
 74607 TGGTCCAAATTCAGAAATCCAGGCTGTGAGGAATCTCATGAAGACCCCT 74558
 339 leu1phe1val1al1le1h1r1cys1al1al1e1gl1met1gl1y1u1ser1gl1up1he1n1 355
 74557 CTCTTTTGGTCACTCACTTGTGCAATCCAGATGGGTGAAGTGAATGCCA 74508
 355 sser1h1s1thr1gl1n1thr1thr1leu1phe1n1s1thr1phe1tyr1asp1leu1leu1leu1 372
 74507 CTCTCACACACAAACAAAGCTGTTCATCTACCTCTATGATCTGTGTGATAC 74458
 372 lu1lys1asn1lys1h1s1h1s1lys1gl1y1val1al1al1aser1asp1he1le1arg 388
 74457 AGAAAAACAACACAAACATAAAGGTGGCTGCAAGTGCATTCATTCGCG 74408
 389 ser1leu1asp1h1s1cys1gl1y1asp1leu1al1al1eul1gl1y1al1phe1ser1h1s1ly 405
 74407 AGCCTGGACACAGTGTGAGACCTAGCTCTGAGGGGTGTGTTCTCCCAA 74358
 405 sp1he1asp1phe1gl1u1leu1gl1n1asp1val1ser1ser1val1asn1gl1asp1val1leu1 422
 74357 GTTGTGATTTGCAACTGCAAGATGTGTCCACGCTGAAGATGAGATGTCTGCG 74308
 422 eut1h1thr1h1r1gl1y1leu1leu1cys1lys1tyr1thr1al1agl1n1arg1phe1lys1pro1lys 438
 74307 TGCACACTGGGCTCTCTCTGTAATATATACACTCAAGGTTCAAGCCCAAAG 74258
 439 tyr1lys1phe1he1h1s1lys1ser1phe1gl1n1gl1y1tr1thr1al1agl1y1n1arg1le 455
 74257 TATTAATTTCTTACAGATCATTCAGAGATACACACAGCAGAGAAAGACT 74208
 455 user1ser1leu1leu1thr1ser1h1s1gl1up1ro1gl1u1val1thr1lys1gl1y1asn1 472
 74207 CACGAGTTTATTGACGCTCATGAGCCAGAGGAGGTGACCAAGGGGAGATG 74158
 472 ly1tyr1leu1gl1n1lys1met1val1ser1le1ser1asp1le1h1r1ser1thr1tyr1ser 488
 74157 GTTACTTGGAGAAAAATGTTTCATTTCCGACATTTACATCCTACTATATAGC 74108

489 Ser1leu1n1arg1tyr1h1n1cys1gl1y1ser1ser1val1gl1u1al1h1r1ar1gal1a 505
 74107 AGCCTGCTCCGGTACCTGTGGGTGATCTGTGGAGCCACACAGGCGTGT 74058
 505 l1met1lys1h1s1leu1al1al1aval1tyr1gl1n1h1s1gl1cys1leu1leu1lys 522
 74057 TATGAACACCTGCGACAGTGTATCAACAGCGCTGCCCTTCTCGACTTTT 74008
 522 er1le1al1al1y1arg1pro1leu1tr1par1gl1n1gl1u1ser1leu1gl1n1ser1val1lys 538
 74007 CCATCGCCCAAGAGGCTCTCTGAGACAGAGAACTTTGCAAAAGTGTAATA 73958
 539 Asn1h1r1thr1gl1n1gl1u1le1leu1lys1al1al1e1asn1le1asn1ser1phe1a 555
 73957 AACACACTGAGCAAGAAATTTGTAAGCCATAAACATCAATTCCTTTGT 73908
 555 lgl1uc1ys1gl1y1le1h1s1leu1tyr1gl1n1gl1u1ser1thr1ser1lys1ser1al1e1us 572
 73907 AGAGTGTGGCATCATTTATATATCAAGAGATACATCCAAATCAGCCCTTGA 73858
 572 er1gl1n1gl1up1he1gl1u1al1phe1he1gl1n1gl1y1s1er1leu1tyr1le1asn1ser 588
 73857 GCCAAGATTTGAAGCTTCTTCTTCAAGGTAAACCTTATATATCAACTCA 73808
 589 gl1y1asn1le1pro1asp1tyr1leu1phe1asp1phe1he1gl1u1h1s1leu1pro1asn1y 605
 73807 GGGAAATCCCCGATTTACTTATTTGACTTCTTTGAAACATTTGGCCCAATGG 73758
 605 sAl1aser1Al1al1e1u1n1s1p1he1le1lys1leu1asp1phe1tyr1gl1y1al1met1a 622
 73757 TGCAGTGGCCCTGGACTTCATTAACCTGAGACTTTTATGGGGGACTATNG 73708
 622 l1aser1tr1p1gl1u1ys1Al1al1agl1u1asp1thr1gl1y1le1h1s1met1gl1u1n 638
 73707 CTTCATGGGAAAGCGTGCAGAAACACAGAGTGAATCCATGGAAGAG 73658
 639 Al1a1pro1gl1u1h1r1tyr1le1pro1ser1arg1Al1al1ser1leu1phe1n1s1r 655
 73657 GCCCAGAAACCTACATTTCCAGCAGGGCTGTATCTTGTGTTCTTCAACTG 73608
 655 p1lys1gl1n1gl1up1he1arg1h1r1leu1gl1u1al1h1r1leu1arg1asp1he1ser1yl 672
 73607 GAAGCAGAGAAATTCAGAGCTCTGAGGTGCACACTCCGGGATTTTCGCAAGT 73558
 672 eu1asn1lys1gl1n1asp1le 677
 73557 TGAATTAAGCAAGATATC 73541
 seq_name: gp_pr:IR2005417
 seq_documentation_block:
 LOCUS IR2005417 1355 bp mRNA PRI 16-JUL-2000
 DEFINITION Homo sapiens mRNA full length insert cdna clone EUROIMAGE 2005417.
 ACCESSION AL389934
 VERSION AL389934.1 GI:9367839
 KEYWORDS F1L_CDNA.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1355)
 AUTHORS Auffray,C., Ansorge,W., Ballabio,A., Estivill,X., Gibson,K.,
 Lehnach,H., Poustka,A. and Lundberg,J.
 TITLE The European IMAGE consortium for integrated Molecular analysis of
 human gene transcripts
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1355)
 AUTHORS Pluvinet,R., Estivill,X., Escarceller,M. and Sumoy,L.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUL-2000) Dept. Genetica Molecular, Institut de
 Recerca Oncologica (IRO), Hospital Duran i Reynals, Av. Gran Via
 s/n Km 2,7 L'Hospitalet de llobregat, 08907 Barcelona, Catalunya,

COMMENT

SPAIN. Tel: ++34-93-260-7775 Fax: ++34-93-260-7776 WWW site:
http://www.iro.es e-mail enquiries: lsumoy@iro.es
BRO-IMAGE Consortium Contact: Aulfray C
CNRS UPR 420 - Genetique Molculaire et Biologie du Developpement
IR 1221 - Rue Guyquet 19, Batiment G - Bp 8
94801 Villejuif Cedex, FRANCE
Tel: ++33-1-49 58 34 98
Fax: ++33-1-49 58 35 09
e-mail: aulfray@infobiogen.fr
This clone is available royalty-free through IMAGE Consortium
distributors.
IMPORTANT: This sequence represents the full insert of this IMAGE
cDNA clone. No attempt has been made to verify whether this
corresponds to the full-length of the original mRNA from which it
was derived.

FEATURES

SOURCE

Location/Qualifiers
1..1355
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="IMAGE cDNA clone 2005417"
/clone_1id="NCI_CGAP_Paul"
/map="2p21-p22"
<1..1321
/codon_start=2
/product="hypothetical protein, weakly similar to
(AF102871) neuronal apoptosis inhibitory protein 2 [Mus
musculus]"
/protein_id="CA897523.1"
/db_xref="GI:9367840"

CDS

BASE COUNT 416 a 265 c 312 g 362 t
ORIGIN
alignment_scores:
Quality: 346.00 Length: 346
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x IR2005417 ..

Align seg 1/1 to: IR2005417 from: 1 to: 1355

679 TyrlenglyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIlely 695
|||||
281 TATCTGGGGAATAATTCAGCTCTGCCACACCTCAGGCTGCAANTAA 330
695 sArgCysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysL 712
|||||
331 GAGATGCTGCTGTGGCTGGAAGCCTCAGTTGGTCTCAGCACCTGTA 380
712 yAsnIleIySerLeuMetValGlnAlaSerProLeuThrIleGluAsp 728
|||||
381 AGAACATTATTCCTCATGTGGGAAGCCACGCCCTCACATAGAACAT 430
729 GluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHisAs 745
|||||
431 GAGAGGACATCATCTGTACAAACCTGAAACCTTAGATTCATATCA 480
745 pLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGlyAsnL 762
|||||
481 CCTACAGATCAACGAGCTCCGGGTGCTGTGACTGACACCTTGGGTAAC 530
762 euLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsnGlnGlu 778

|||||
531 TGAGAACCTTACAAAGCTTCATTAAGATTAACATTAAGATGAAGA 580
779 AspAlaIleLysLeuAlaGlyLeuLysAsnLeuLysLysMetCysLe 795
581 GATGCTATTAACCTAGAGGCTGGAAGAACTTGAAGAGTGTGT 630
795 uPheHisLeuThrHisLeuSerAspIleGlyGlyLysMetAspTyrIle 812
|||||
631 ATTTCATTGACCACTGCTGTGACATTTGAGAGGAAATGATTAATG 680
812 alLysSerLeuSerSerGluProCysAspLeuGluIleGlnLeuVal 828
|||||
681 TCAAGCTCTGTCAAGTGAACCTGTGACCTGAAATAATTAATGATG 730
829 SerCysCysLeuSerSerAlaAsnAlaValLysIleLeuAlaGlnAsn 845
731 TCCTGCTGCTTGTCTGCANAATGCAGTGAATAATCTAGCAGATCTTA 780
845 sAsnLeuValLysLeuSerIleLeuAspLeuSerGluAsnTyrLeuGlu 862
|||||
781 CATTTGGTCAAACTGAGCATCTTATTAACAGAAATTAACCTGGA 830
862 yAspGlyAsnGlnAlaLeuHisGluLeuIleAspArgMetAsnValLe 878
|||||
831 AAGATGGAATGAAGCTTTCATGAACATGATGACAGATGAACGTGTA 880
879 GluGlnLeuThrAlaLeuMetLeuProTyrGlyCysAspValGlnGly 895
881 GAAACGCTACCGCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCA 930
895 rLeuSerSerLeuLeuLysHisLeuGlnGluValProGlnLeuValLys 912
|||||
931 CCTGACACACCTGTTGAACATTTGGAGAGAGGCCACAACCTGTCAC 980
912 euGlyLeuLysAsnThrPargLeuThrAspThrGluIleArgGlyLeu 928
|||||
981 TTGGGTTGAAAACTGGAGACTCACAGATACAGATTAAGAAATTAAG 1030
929 AlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeu 945
|||||
1031 GCATTTTGGAAAGAACCTCTGAAATACTTCACACAGTGAATTTGG 1080
945 acLysAsnArgValSerSerAspGlyTyrPheAlaPheMetGlyValPhe 962
1081 GGGAAATCGTGTGAGCAGATGATGATGCTGCTGCTGCTGCTGCTG 1130
962 LuAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPheLe 978
|||||
1131 AGAATCTTAAGCAATTAAGTCTTTTGGACTTAACTTAAGCAATTTCTA 1180
979 ProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeu 995
|||||
1181 CCTGATCCAGCATTAAGTCAAGAAACCTTACCAAGTTCATCAAGTTA 1230
995 rPheLeuGlnGlnAlaArgLeuValGlyTyrPheAspAspAspAsp 1012
|||||
1231 TTTTTCGAAAGAGCTGAGCTGTTGGTGGCAATTTGATGATGATG 1280
1012 euSerValIleThrGlyAlaPheLysLeuValThrAla 1024
|||||
1281 TCAGTGTATTACAGGTGCTTTAACTAGTAATGCT 1318

seq_name: gb_pr:AY027788

seq_documentation_block:

LOCUS AY027788 mRNA

DEFINITION Homo sapiens CLANB (CLAN1) mRNA, complete cds.

ACCESSION AY027788

VERSION AY027788.1 GI:14324114

KEYWORDS human.

SOURCE ORGANISM Homo sapiens

20-JUL-2001

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1395)

REFERENCE 1 (bases 1 to 1395)

AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.

JOURNAL Clin., a novel human ced-4-like gene

MEDLINE 21365712

PUBMED 11472070

REFERENCE 2 (bases 1 to 1395)

AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES

source Location/Qualifiers

1..1395

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2p22-P21"

/tissue_type="lung"

1..1395

/gene="CLAN1"

277..1356

/gene="CLAN1"

/codon_start=1

/product="CLANB"

/protein_id="AAK14777.1"

/db_xref="GI:14324115"

/translation="MNFIDNSRALIQRMGTIVYIKQITDLEFVNVNLNREEVNIIICCE KVEODPARGIIIMILKKGESCNLFKSLKEWNYPLEQDLNGSLTSLGNLKNLTK LIMDNIKNEEDAKIKLAGKMLKKMLFHLTHLSPIGEMDIYKLSLSEPCDLEEI OLVSCLSANAKVILIAONLHNLVKSIIIDLSNVYLEKDNELHELDIMNVYLOLTA LMLPGCCVOGSLSLKHLKEEYPOLYVKIGLKNMRTDPEIRILGAFRKNPLKNTQO LMLAGNRSSDGNLAFMGVEFNKQLVFFDFSTKEFLPDPALVKLSQVLSKLTFLOE ARLVQWQFDDDLISVITGAFKLVTAT"

BASE COUNT 436 a 248 c 327 g 384 t

ORIGIN

alignment_scores:

Quality: 271.00 Length: 271

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AY027788 ..

Align seq 1/1 to: AY027788 from: 1 to: 1395

754 GlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMe 770

|||||

541 GGTTGACTGACAGCTGGTGAAGAACTTACAAAGCTCATATAT 590

|||||

770 TASPASNILEYMETASNGIUGIUAAPALAIIEYSLAUAGIUGIYL 787

|||||

591 GGATTAACATAAAGATGATGAAGAAGATGCTATAAATACCTGAAGCC 640

|||||

787 EULYSANLEULYSLEMETCYSLEUPHEHISLEUTHRHSLEUSERASP 803

|||||

641 TGAAGAAACCTGAAGAAGATGTTTATTTCAATTGACCCACTTGTCTGAC 690

|||||

804 ILEGLIUGIUGIUMETASPTVRIIEVALYSEUSERSERGIUPROCY 820

|||||

691 ATTGGAGAGGGAATGATTAATGCTCAAGTCTCTGTCAGTGAACCTG 740

|||||

820 SASPLEUGIUGIULEGINLEUVALSERCYSCYSLUSERALASNALAY 837

|||||

741 TGACCTTGAGAGAAATTCATTAATGCTCCTGCTGCTGCTCAATGAG 790

|||||

837 ALLYSILEUVALAGIASNLEUHSANLEUVALYSLUSERILEU 853

|||||

791 TGAAGAACTCTAGCTCAGAAATCTTCACAAATTTGTCAAACTGACATTCCT 840

854 ASPLEUSERGIUASNTRYILEUGIULYSASPGIYASNGIUALALEUHSIGL 870

|||||

841 GATTTATCAGAAAATTTACCTGGAAAAAGATGAAATGAACCTCTTACAGA 890

|||||

870 ULEULIeAspArgMeTasNValLeuGluGlnLeuThrAlaLeuMetLeuP 887

|||||

891 ACTGATCCACAGAGATGACGTGCTAGAACAGCTACCCGACGATGATGCTGC 940

|||||

887 ROTTPGLCYASAPVALGNGIYSERLEUSERSERLEULEULYSHISLEU 903

|||||

941 CCTGGGCTGTGACCTGCACAGCACCTGACAGAGCTCTTGGAACATTTG 990

|||||

904 GIUGIUALPROGLINLEUVALYSLLEUGIYLEULYSANTRPARLEUTH 920

|||||

991 GAGGAGGTCCACACACTGCTCAAGCTTGCGTTGAAAACCTGGAGACTCAC 1040

|||||

920 RASPThrGluIleArgIleLeuGlyAlaPhePheGlyLysAsnProLeuL 937

|||||

1041 AGATACAGAGATTGAATTTAGTGCATTTTGTGAAAAGAACCTCTGCA 1090

|||||

937 YSASNPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAspCly 953

|||||

1091 AAAACTTCACAGAGTTGATTTGCGGGAATTCGTGAGCAGTGATGCA 1140

|||||

954 TRPLEUALIAPHEMETGLYVALPHEGLUASNLEULYSGLINLEUVALPHEPH 970

|||||

1141 TGGCTTGCTTCATCGGTGTATTGTAGAAATCTTAAGCAATTAAGTGTGTTT 1190

|||||

970 EASPPheserThrLysGluPheLeuProAspProAlaLeuValArgLysL 987

|||||

1191 TGACTTTAGTACTAAGAATTTCTACTGCATGCACACATTAAGCAGAAAAC 1240

|||||

987 EUSERGlnValLeuSerLysLeuThrPheLeuGlnGlnAlaArgLeuVal 1003

|||||

1241 TTAGCAAGGTGTATACCAAGTTACTTTCTGCAGAGAGCTAGAGCTTGT 1290

|||||

1004 GLYTRPGINPheAspAspAspLeuSerValIleThrGlyAlaPheLys 1020

|||||

1291 GCGTGCAATTTGATGATGATGATCTCAGTGTATTATACAGGTGCTTTTAA 1340

|||||

1020 SLEUVALThrAla 1024

|||||

1341 ACTAGTAACTGCT 1353

seq_name: gb-PR:AY027789

seq_documentation_block:

LOCUS AY027789 768 bp mRNA PRI 20-JUL-2001

DEFINITION Homo sapiens CLAN1 (CLAN1) mRNA, complete cds.

ACCESSION AY027789

VERSION AY027789.1 GI:14324116

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 768)

REFERENCE 1 (bases 1 to 768)

AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.

JOURNAL Clin., a novel human ced-4-like gene

MEDLINE 21365712

PUBMED 11472070

REFERENCE 2 (bases 1 to 768)

AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES

source Location/Qualifiers

1..768

/organism="Homo sapiens"

/db_xref="taxon:9606"

```

gene      /chromosome="2"
          /map="2p22-p21"
          /tissue_type="lung"
          1..768
CDS       /gene="CLAN1"
          277..747
          /codon_start=1
          /product="CLAN1"
          /protein_id="AAK14778.1"
          /db_xref="GI:14324117"
          /translation="MNFIKDNRRALIQRMGWIVIKQIDDLFVWVLRNEFNITICE
          KYEDDAARGITIMILKSGSESCNLEFLSKLEKNYTFLODLNCSLPFQTSSEDLDDLA
          QDLKDLHTPSPFLNYPYIGEDIDIIIFNLSKTFTEPVLWKDQHHRRVQLTLVL"

```

```

BASE COUNT      218 a      157 c      180 g      213 t
ORIGIN

```

```

alignment_scores:
  Quality: 154.00      Length: 154
  Ratio: 1.000          Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-09-697-089-2 x AY027789 ..

```

```

Align seg 1/1 to: AY027789 from: 1 to: 768

```

```

1 MetAsnPhelIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyme 17
|||||
277 ATCAATTTCTTAAGACATACCGCGCCCTTATTCAAGATGGGAAT 326
|||||
17 LThrValIleLysGlnIleThrAspSpleuPheValITrpAsnValLeuA 34
|||||
327 GACTGTATTAAGCAAAATCACAGATGACCTATTGTATGCAATGCTCTGA 376
|||||
34 snArgGluGluValAsnIleIleCysGlyValValGluGlnAspAla 50
|||||
377 ATGCGCAAGAAAGTAACATCATTTGCTGCGAAGAGTGAGCAGAGATGCT 426
|||||
51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGlySerCysAs 67
|||||
427 GCTAGAGGATCATTCACATGATTTGAAAAGGTTCAAGATCTCTTAA 476
|||||
67 nLeuPheLeuLysSerLeuLysGluITrpAsnITyrProLeuPheGlnAsp 84
|||||
477 CCTCTTTCTTAATCCCTTAAGAGTGAGACTATCTCTATTTCAGACT 526
|||||
84 euAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
|||||
527 TGAATGACAAAGCTTTTTCATCAGACATCAGAAAGAGACTTGGAGAT 576
|||||
101 LeuAlaGlnAspLeuLysAspLeuTyrHisIThrProSerPheLeuAsnPh 117
|||||
577 TTGGCTCAGAGATTTAAAGACTTGTACCATACCCCATCTTTTCTGAACTT 626
|||||
117 eTyrProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerIThrP 134
|||||
627 TTATCCCTTGGTGAAGATATGACATATTATTTAACTTGAAAAGCACT 676
|||||
134 heThrGluProValLeuITrpArgLysAspGlnHisIThrHisArgValGlu 150
|||||
677 TCACAGAACCTGTCTGTGAGAGAACACACCATCATCCCGCTGGAG 726
|||||
151 GlnLeuIThrLeu 154
|||||
727 CACCTACCCCTA 738

```

```

seq_name: gb_pr:CNS01DS8

```

```

seq_documentation_block:
LOCUS      CNS01DS8      162692 bp      DNA      PRI      02-MAR-2000
DEFINITION BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 563N04

```

```

of RPCI-11 library from chromosome 2 of Homo sapiens (Human),
complete sequence.
ACCESSION  AL121658
VERSION    AL121658.2  GI:7159619
KEYWORDS   HTG: HTCS_DRAFT; SPG4 genomic DNA interval.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 162692)
Hazen,J., Fomnechten,N., Mavel,D., Patenotte,C., Samson,D.,
Artiguenave,F., Davoine,C.S., Cruaud,C., Durf,A., Wincker,P.,
Brotlier,P., Catolico,L., Barbe,V., Burgunder,J.M.,
Prud'Homme,J.F., Brice,A., Fontaine,B., Hellig,R. and
Weissenbach,J.

```

```

TITLE      Spastin, a novel AAA protein, is altered in the most frequent form
of autosomal dominant spastic paraplegia
JOURNAL    Nat. Genet. (1999) In press
REFERENCE  2 (bases 1 to 162692)
AUTHORS    Genoscope.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (29-FEB-2000) to the EMBL/GenBank/DBJ databases
COMMENT    On Mar 6, 2000 this sequence version replaced gi:6002391.
FEATURES   location/Qualifiers
SOURCE     1..162692
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="2"
           /clone_lib="RPCI-11"
           /clone="563N04"

```

```

BASE COUNT      44981 a      32017 c      32887 g      52792 t      15 others
ORIGIN

```

```

alignment_scores:
  Quality: 97.00      Length: 97
  Ratio: 1.000          Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-09-697-089-2 x CNS01DS8/rev ..

```

```

Align seg 1/1 to reverse of: CNS01DS8 from: 1 to: 162692

```

```

928 GtAlaPhePheGlyLysAsnProLeuLysAsnPhelGlnLeuAsnLe 944
|||||
152217 GGTGCATTTTGTGAAGAAACCCCTGAAAACTTCAGCAGTGTGATTT 152168
|||||
944 uAlaGlyAsnArgValSerSerAspGlyITrpLeuAlaPheMetGlyValP 961
|||||
152167 GCGCGGAAATCGTGTGAGCATGATGATGGCTTGCCTTCATGCGGTGAT 152118
|||||
961 heGluAsnLeuLysGlnLeuValPhePheAspPheSerIThrLysGluPhe 977
|||||
152117 TTGGAATCTTAAGCAATGTAGTGTTTTTCACCTTACTTAAGAAATTT 152068
|||||
978 LeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerITrLys 994
|||||
152067 CTACCTGATCCAGCATATAGTCAGAAACTTAGCAACTGTATCCAGACTT 152018
|||||
994 uThrPheLeuGlnGluAlaArgLeuValGlyITrpGlnPheAspAspAspA 1011
|||||
152017 AACCTTCTCAAGAAAGCTAGCGCTTGTGGTGCAATTTGATGATGATG 151968
|||||
1011 spleuSerValIleIThrGlyAlaPheLysLeuValITrAla 1024
|||||
151967 ATCTCAGTGTATTACAGGTGCTTTTAACACTAGTAAGTACTGCT 151927

```

```

seq_name: gb_htg:AC011232

```

```

seq_documentation_block:
LOCUS      AC011232      185281 bp      DNA      HTG      10-MAR-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,

```

```

7 unordered pieces.
AC011232
AC011232.7 GI:13270720
HTG: HTGS_PHASE1, HTGS_DRAFT, HTGS_FULLTOP.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 185281)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 185281)
Waterston,R.H.
Direct Submission
Submitted (04-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 10, 2001 this sequence version replaced g1:9799811.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0078E13
----- Summary Statistics -----
Sequencing vector: M13, 578
Sequencing vector: plasmid, 408
Chemistry: Dye-primer EM, 48% of reads
Chemistry: Dye-terminator Big Dye, 50% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182798 bases at least Q40
Consensus quality: 183538 bases at least Q30
Consensus quality: 184045 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 184681; sum-of-ctags
Quality coverage: 7.07 in Q20 bases; sum-of-ctags
Quality coverage: 7.33 in Q20 bases; sum-of-ctags
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1985: contig of 1985 bp in length
1986
2085: gap of unknown length
2086
8183: contig of 6098 bp in length
8184
8283: gap of unknown length
8284
22741: contig of 14458 bp in length
22742
22841: gap of unknown length
22842
41446: contig of 18605 bp in length
41447
41546: gap of unknown length
41547
69282: contig of 27636 bp in length
69283
11231: contig of 42849 bp in length
11232
185281: contig of 73050 bp in length.
11233
Location/Qualifiers
1..185281
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-78E13"
1..1985
/note="assembly_name:Contig11"
misc_feature
2086..8183
/note="assembly_name:Contig12"
misc_feature
8284..22741
/note="assembly_name:Contig13"
FEATURES
source

```

```

misc_feature 22842..41446
/note="assembly_name:Contig14"
misc_feature 41547..69283
/note="assembly_name:Contig15"
misc_feature 69283..112131
/note="assembly_name:Contig16"
misc_feature 112132..185281
/note="assembly_name:Contig17"
BASE COUNT 52719 a 36957 c 37187 g 57816 t 602 others
ORIGIN

alignment_scores:
Quality: 97.00 Length: 97
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-697-089-2 x AC011232/rev ..

Align seg 1/1 to reverse of: AC011232 from: 1 to: 185281

928 GYLAAPHEPHEGLYLYSAsnProLeuLysAsnPhenGlnLeuAsnLe 944
|||||
160182 GGTGCATTTTGGAAAGAACCCCTCGAAAACTCCAGCAGTGAATTT 160133
|||||
944 WALGILYSnArGVALSerSerAspGLYTrpLeuAlaPheMetGLYValP 961
|||||
160132 GCGCGGAATGCTGTGACAGATGATGATGCTTCCTTCAGGGTGTAT 160083
|||||
961 heGLUAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPhe 977
|||||
160082 TTGACAACCTTAGCAATTAAGTGTGTTTGGACTTAGTACACTTAAGAATTT 160033
|||||
978 LeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLe 994
|||||
160032 CTACCTGATCCAGCATTAAGTCAGAAACTTAGCCAAAGTGTATCCAACTT 159983
|||||
994 uThrPheLeuGlnGlnAlaArgLeuValGLYTrpGlnPheAspAspAspA 1011
|||||
159982 AACTTTCTTGCAAGAAAGCTAGCGCTGTGGGTGGCAATTTGATGATG 159933
|||||
1011 sPLeserValIleThrGlyAlaPheLysLeuValThAlA 1024
|||||
159932 ATTCACAGTGTATTACAGGCTCTTTAACTAGTACTGCT 159892
|||||

seq_name: gb_pr:AY027790

seq_documentation_block:
LOCUS AY027790 578 bp mRNA PRI 20-JUL-2001
DEFINITION Homo sapiens CLAND (CLAN1) mRNA, complete cds.
ACCESSION AY027790
VERSION AY027790.1 GI:14324118
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 578)
Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE
Clan, a novel human ced-4-like gene
JOURNAL
Genomics. 75 (1-3), 77-83 (2001)
MEDLINE
21365712
PUBMED
11472070
REFERENCE
2 (bases 1 to 578)
AUTHORS
Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE
Direct Submission
JOURNAL
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
Location/Qualifiers
1..578
/organism="Homo sapiens"
FEATURES
source

```

/db_xref="taxon:9606"
 /chromosome="2"
 /map="2p22-p21"
 /tissue_type="lung"
 1..578
 /gene="CLAN1"
 /gene="CLAN1"
 277..555
 /note="CARD protein"
 /codon_start=1
 /product="CLAND"
 /protein_id="AAK14779.1"
 /db_xref="GI:14324119"
 /translation="MFKIDNSRALIQMGKTVIKQITDILFVWNVINREEVNIICCE
 KVEDAARGIIMHLKSGESCNLFKSLKEMNYPLEFDLNGOSILTA"

BASE COUNT 172 a 106 c 143 g 157 t
 ORIGIN

alignment_scores:
 Quality: 89.00 Length: 89
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AY027790 ..

Align seg 1/1 to: AY027790 from: 1 to: 578

1 MetAspHelelelyspasnserrargAlaleuilegluargmetglyme 17
 ||||||||||||||||||||||||||||||||||||||||||||||||
 277 ATGATATTCATTAAGACACATACGCCCTTATTCATAAGATGGGAAT 326
 17 tthrValilelysglnlethrAspAspleupheValtrpAsnValleua 34
 ||||||||||||||||||||||||||||||||||||||||||||||||
 327 GACTGTTATAAGCAATCACAGATGACCTATTGTATGSAATGTTCTGA 376
 34 snArggluGlulvalasnleileCysCysglulysValgluInaspala 50
 ||||||||||||||||||||||||||||||||||||||||||||||||
 377 ATCGCGAAGAGTAAACATCATTTGCTCGAGAGGTGGAGCATGCT 426
 51 AlaArgGlylelleHleHMetleleuLylysglySerclnserCysas 67
 ||||||||||||||||||||||||||||||||||||||||||||||||
 427 GCTAGAGGATCATTCACATGATTTGAAAAAGGTTCCAGAGCTCTGTAA 476
 67 nleupHeleuLySerleuLyglultrpAsnTyrrProleupheGlnaspl 84
 ||||||||||||||||||||||||||||||||||||||||||||||||
 477 CCTCTTCTTAATCCCTTAAGAGAGTGAACATATCTCTATTTCAGAGACT 526
 84 euAsnGlylnserLeu 89
 ||||||||||||||||||||||||||||||||||||||||||||||||
 527 TGAATGGACAAAGTCTT 543

seq_name: gb_sts:G55568

seq_documentation_block:

LOCUS G55568 553 bp DNA STS 30-MAR-2000
 DEFINITION SHGC-100923 Human Homo sapiens STS genomic, sequence tagged site.
 ACCESSION G55568
 VERSION G55568.1 GI:6120887
 KEYWORDS STS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 553)
 AUTHORS Olivier,M. and Cox,D.R.
 TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
 JOURNAL Unpublished (2000)
 COMMENT

Contact: Michael Olivier, David R. Cox
 Stanford Human Genome Center
 Stanford University School of Medicine

4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
 Tel: (650) 320-5800
 Fax: (650) 320-5801
 Email: olivier@shgc.stanford.edu
 Primer A: AATAGAGGGGCAAAATAGCAAA
 Primer B: TAACACCCCTTTGTCTCTCA
 STS size: 322
 PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
 Denaturation: 94 degrees C for 30 seconds
 Annealing: 60 degrees C for 30 seconds
 Polymerization: 72 degrees C for 23 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9700

Protocol:
 Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 AmpliTaq Gold Polymerase: 0.07 units/ul
 Total Vol: 5 ul

Buffer:
 MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES
 source
 Location/Qualifiers
 1..553
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="2"
 /clone_lib="Human"

STS
 primer_bind 5..27
 primer_bind complement(304..326)
 BASE COUNT 170 a 107 c 114 g 162 t
 ORIGIN

alignment_scores:
 Quality: 56.00 Length: 56
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x G55568/rev ..

Align seg 1/1 to reverse of: G55568 from: 1 to: 553

873 AspArgMetAsnValleuGlulnleuThrAlaleuMetleuProtrpgl 889
 ||||||||||||||||||||||||||||||||||||||||||||||||
 256 GACAGAGTGAACGTGCTAGACACAGCTCACCGCACTGATGCTGCCCTGGGG 207
 889 YCysaspValglnglySerleuSerleuLeuLyshlsleuGluluv 906
 ||||||||||||||||||||||||||||||||||||||||||||||||
 206 CTGTGACGTGCAGGCGAGCCTGAGCAGCCCTGTAACATTTGGAGAGG 157
 906 alProGlnleuVallyleuGlyleuLyAsntrpArgleuThrAspThr 922
 ||||||||||||||||||||||||||||||||||||||||||||||||
 156 TCCACACACTCGTCAAGCTTGCGTTGAAAAAAGTGAAGACTCACAGATACA 107
 923 GluileArglleuGly 928
 ||||||||||||||||||||||||||||||||||||||||||||||||
 106 GAGATTGAGATTITAGGT 89

seq_name: gb_pr:AP000783

seq_documentation_block:

LOCUS AP000783 114233 bp DNA PRI 30-MAY-2001
 DEFINITION Homo sapiens genomic DNA, chromosome 11q clone: CMB9-12D18, complete

sequence.
 AP000783
 VERSION AP000783.4 GI:14245745
 KEYWORDS HTG.
 SOURCE Homo sapiens DNA, clone CMB9-12D18.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Homo sapiens genomic DNA
 JOURNAL Published Only in Database (1999) In press
 AUTHORS 2 (bases 1 to 114233)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 COMMENT On May 29, 2001 this sequence version replaced gi:9844969.
 FEATURES
 source
 1..114233
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="11q"
 /clone="CMB9-12D18"
 /map="11q"
 /clone="CMB9-12D18"
 BASE COUNT 30375 a 24634 c 25670 g 33554 t
 ORIGIN
 alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x AP000783/rev ..
 Align seg 1/1 to reverse of: AP000783 from: 1 to: 114233
 452 G|YAARG|euser|Ser|Leu|Thr|Ser 461
 ||||||||||||||||||||||||||||
 75996 GGCAGGCGACTGCTCTCTTAACCTCT 75967
 seq_name: gb_htg:AC080039
 seq_documentation_block:
 LOCUS AC080039 158375 bp DNA HTG 07-DEC-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-398M20 map 11, WORKING DRAFT
 SEQUENCE 5 unoriented pieces.
 AC080039
 AC080039.2 GI:11597030
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 158375)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 11, clone RP11-398M20
 JOURNAL Unpublished
 2 (bases 1 to 158375)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Chesepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Deaellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,

Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Laroque,K., Lamazares,R., Lander,T., Lehotzky,J., Levine,R., Lieu,C., Liu,G., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menus,L., Milnova,T., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanl,C., Pollara,Y., Raymond,C., Rebbeck,M., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Sounez,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,D., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 TITLE Direct Submission
 JOURNAL Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Dec 7, 2000 this sequence version replaced gi:10280794.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11080
 Center clone name: 398_M_20
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 155266 bases at least Q40
 Consensus quality: 157049 bases at least Q30
 Consensus quality: 157565 bases at least Q20
 Insert size: 163000; agarose-fp
 Insert size: 157975; sum-of-ctrls
 Quality coverage: 7.1 in Q20 bases; agarose-fp
 Quality coverage: 7.3 in Q20 bases; sum-of-ctrls

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 52270: contig of 52270 bp in length
 * 52271 52370: gap of 100 bp
 * 52371 56080: contig of 3710 bp in length
 * 56081 56180: gap of 100 bp
 * 56181 94106: contig of 37926 bp in length
 * 94107 94206: gap of 100 bp
 * 94207 121121: contig of 26915 bp in length
 * 121122 121221: gap of 100 bp
 * 121222 158375: contig of 37154 bp in length.
 FEATURES
 source
 1..158375
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-398M20"
 /clone_11b="RP11-11 Human Male BAC"
 1..52270
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 52371..56080
 /note="assembly_fragment"
 56181..94106
 /note="assembly_fragment"

```

misc_feature      94207..121121
                  /note="assembly-fragment"
misc_feature      112122..158375
                  /note="assembly-fragment"
                  clone_end:T7
                  vector_side:right"
BASE COUNT      45378 a 36237 c 35136 g 41223 t 401 others
ORIGIN
alignment_scores:
    Quality:      10.00      Length:      10
    Ratio:         1.000      Gaps:         0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
    US-09-697-089-2 x AC080039 ..
Align seg 1/1 to: AC080039 from: 1 to: 158375
452 GlyAArgLeuSerSerLeuThrSer 461
|||||
55769 GGCAGCGACGTGCTCTGTACCTCT 55798
seq_name: qb_htg:AC036227
seq_documentation_block:
LOCUS      AC036227 168346 bp DNA HTG 12-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-593K21 map 11, WORKING DRAFT
ACCESSION      AC036227
VERSION      AC036227.2 GI:7770605
KEYWORDS
SOURCE      HTGS_PHASE1; HTGS_DRAFT.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 168346)
  Birren, B., Linton, L., Nusbaum, C. and Lander, E.
  Homo sapiens chromosome 11, clone RP11-593K21
  Unpublished
  2 (bases 1 to 168346)
  Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
  Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
  Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
  Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
  Collymore, C., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
  Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
  Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
  Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
  Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karakas, A.,
  Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
  Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
  McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
  McElrath, J., Menus, L., Mihova, T., Miranda, C., Mlenda, V., Morrow, J.,
  Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
  O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
  Pisanic, C., Polara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
  Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
  Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
  Testafae, S., Theodore, J., Tirelli, A., Travers, M., Tridillo, J.,
  Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
  Young, G., Zainoun, T., Zimmer, A. and Zody, M.
  Direct Submission
  Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On May 12, 2000 this sequence version replaced gi:753898.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: W1BR

```

```

Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 19442
Center clone name: 593_K21
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152315 bases at least 940
Consensus quality: 160567 bases at least 930
Consensus quality: 164240 bases at least 920
Insert size: 174000; agarose-fp
Insert size: 166346; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      1184: contig of 1184 bp in length
*      1185 1284: gap of 100 bp
*      1285 2630: contig of 1346 bp in length
*      2631 2730: gap of 100 bp
*      2731 4080: contig of 1350 bp in length
*      4081 4180: gap of 100 bp
*      4181 5649: contig of 1469 bp in length
*      5650 5749: gap of 100 bp
*      5750 8699: contig of 2950 bp in length
*      8700 8799: gap of 100 bp
*      8800 11212: contig of 2413 bp in length
*      11213 11312: gap of 100 bp
*      11313 15075: contig of 3763 bp in length
*      15076 15175: gap of 100 bp
*      15176 19034: contig of 3859 bp in length
*      19035 19134: gap of 100 bp
*      19135 24110: contig of 4976 bp in length
*      24111 24210: gap of 100 bp
*      24211 28836: contig of 4626 bp in length
*      28837 28936: gap of 100 bp
*      28937 35471: contig of 6535 bp in length
*      35472 35571: gap of 100 bp
*      35572 42357: contig of 6786 bp in length
*      42358 42457: gap of 100 bp
*      42458 48426: contig of 5569 bp in length
*      48427 48526: gap of 100 bp
*      48527 55409: contig of 6883 bp in length
*      55410 55509: gap of 100 bp
*      55510 60756: contig of 5247 bp in length
*      60757 60856: gap of 100 bp
*      60857 70638: contig of 9782 bp in length
*      70639 70738: gap of 100 bp
*      70739 84684: contig of 13946 bp in length
*      84685 84784: gap of 100 bp
*      84785 101050: contig of 16266 bp in length
*      101051 101150: gap of 100 bp
*      101151 122254: contig of 21104 bp in length
*      122255 122354: gap of 100 bp
*      122355 145955: contig of 23601 bp in length
*      145956 146055: gap of 100 bp
*      146056 168346: contig of 22291 bp in length.
FEATURES
Source
1..168346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-593K21"

```

```

misc_feature /clone.lib="RPCI-11 Human Male BAC"
1. .1184
misc_feature /note="assembly-fragment"
1285. .2630
misc_feature /note="assembly-fragment"
2731. .4080
misc_feature /note="assembly-fragment"
4181. .5649
misc_feature /note="assembly-fragment"
5750. .8699
misc_feature /note="assembly-fragment"
8800. .11212
misc_feature /note="assembly-fragment"
11313. .15075
misc_feature /note="assembly-fragment"
15176. .19034
misc_feature /note="assembly-fragment"
19135. .24110
misc_feature /note="assembly-fragment"
24211. .28836
misc_feature /note="assembly-fragment"
clone_end:T7
vector_side:right"
28937. .35471
misc_feature /note="assembly-fragment"
35572. .42357
misc_feature /note="assembly-fragment"
42458. .48426
misc_feature /note="assembly-fragment"
48527. .55409
misc_feature /note="assembly-fragment"
55510. .60756
misc_feature /note="assembly-fragment"
clone_end:SP6
vector_side:right"
60857. .70638
misc_feature /note="assembly-fragment"
70739. .84684
misc_feature /note="assembly-fragment"
84785. .101050
misc_feature /note="assembly-fragment"
101151. .122254
misc_feature /note="assembly-fragment"
122355. .145955
misc_feature /note="assembly-fragment"
146056. .168346
misc_feature /note="assembly-fragment"
BASE COUNT 51423 a 33737 c 33199 g 47987 t 2000 others
ORIGIN

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AC036227 ..
Align seg 1/1 to: AC036227 from: 1 to: 168346
988 SerginValleuserLyLeuThrPhelu 997
|||||
56164 TCACAGGCTTGAGCAGCTAACCTTCTG 56193

seq_name: gb_htg:AC087661

seq_documentation_block:
LOCUS AC087661 176593 bp DNA HTG 23-MAY-2001
DEFINITION Homo sapiens chromosome 11 clone RP5-1002E13 map 11, WORKING DRAFT
SEQUENCE 5 unordered pieces.
ACCESSION AC087661
VERSION AC087661.2 GI:14190709

```

```

KEYWORDS HTGS_PHASE1, HTGS_DRAFT, HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 176593)
AUTHORS Birren,B., Linton,L., Nusbaum,C., and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP5-1002E13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176593)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Develiano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,W., Grand-Pierre,N.,
Hagos,B., Heathord,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Labocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,D., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
Sougniez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Sudramanlan,A., Talamas,J., Teste,J., Theodore,J.,
Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 23, 2001 this sequence version replaced 1:12229436.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 112140
Center clone name: 1002_E13
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175018 bases at least Q40
Consensus quality: 175722 bases at least Q30
Consensus quality: 176018 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 176193; sum-of-ctrls
Quality coverage: 9.2 in Q20 bases; agarose-fp
Quality coverage: 9.2 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1100: contig of 1100 bp in length
* 1101 1200: gap of 100 bp
* 1201 18982: contig of 17782 bp in length
* 18983 19082: gap of 100 bp
* 19083 48520: contig of 29438 bp in length
* 48521 48620: gap of 100 bp
* 48621 86819: contig of 38199 bp in length

```

```

*      86820 86919: gap of 100 bp
*      86920 176593: contig of 89674 bp in length.
FEATURES
  source
    1..176593
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="11"
      /map="11"
      /clone="RP5-1002E13"
      /clone_lib="RPCI Human PAC library 5"
      1..1100
      /note="assembly_fragment"
      /clone_end:SP6
      vector_side:left"
      1201..18982
      /note="assembly_fragment"
      19083..48520
      /note="assembly_fragment"
      48621..86819
      /note="assembly_fragment"
      86920..176593
      /note="assembly_fragment"
BASE COUNT   54490 a 34484 c 34383 g 52836 t   400 others
ORIGIN
alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
  US-09-697-089-2 x AC087661 ..
Align seg 1/1 to: AC087661 from: 1 to: 176593
988 SergInValleuSerlystLeuthrPheleu 997
|||||
164712 TCACAGGCTCTGACGAGTTAAGCTTCTCG 164741
seq_name: gb_hlg:AC011006
seq_documentation_block:
  LOCUS      AC011006      179901 bp      DNA      HTG      24-JAN-2001
  DEFINITION Homo sapiens chromosome 11 clone RP11-31123 map 11, WORKING DRAFT
  SEQUENCE   4 unordered pieces.
  AC011006
  AC011006 4 GT:12408507
  VERSION   HTG: HTGS_PHASE1; HTGS_DRAFT.
  KEYWORDS  human.
  SOURCE     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 179901)
      Birten,B., Linton,L., Nusbaum,C. and Lander,E.
      Homo sapiens chromosome 11, clone RP11-31123
      Unpublished
    2 (bases 1 to 179901)
      Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
      Baldwin,J., Barne,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
      Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
      Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
      Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
      Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
      Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
      Lebecky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
      McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
      Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
      Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
      Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
      Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
      Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL
  AUTHORS
  REFERENCE
  AUTHORS

```

```

JOURNAL
  Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Jan 24, 2001 this sequence version replaced gi:8096829.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIRB
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: 31_123
  Center clone name: 31_123
  ----- Summary Statistics
  Sequencing vector: M13: M77815; 47% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.960731
  Consensus quality: 178118 bases at least Q40
  Consensus quality: 178909 bases at least Q30
  Consensus quality: 179215 bases at least Q20
  Insert size: 176000; agarose-fp
  Insert size: 179601; sum-of-contigs
  Quality coverage: 10.2 in Q20 bases; agarose-fp
  Quality coverage: 10.0 in Q20.
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 4 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 1208: contig of 1208 bp in length
  * 1209 1308: gap of 100 bp
  * 1309 2363: contig of 1055 bp in length
  * 2364 2463: gap of 100 bp
  * 2464 3567: contig of 1104 bp in length
  * 3568 3667: gap of 100 bp
  * 3668 179901: contig of 176234 bp in length.
FEATURES
  source
    1..179901
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="11"
      /map="11"
      /clone="RP11-31123"
      /clone_lib="RPCI-11 Human Male BAC"
      1..1208
      /note="assembly_fragment"
      1309..2363
      /note="assembly_fragment"
      2464..3567
      /note="assembly_fragment"
      3668..179901
      /note="assembly_fragment"
BASE COUNT   54050 a 35987 c 35567 g 53997 t   300 others
ORIGIN
alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
  US-09-697-089-2 x AC011006/rev ..
Align seg 1/1 to reverse of: AC011006 from: 1 to: 179901
988 SergInValleuSerlystLeuthrPheleu 997
|||||

```

83785 TCACAGGCTTGACCAAGTTAACCTTCTG 83756
seq_name: gb_sts:A0048829

seq_documentation_block: 222 bp DNA STS 20-JAN-2000
LOCUS A0048829
DEFINITION Rattus norvegicus, OTSUKA clone, 773h04, microsatellite sequence,
sequence tagged site.
ACCESSION A0048829
VERSION A0048829.1 GI:6722000
KEYWORDS STS.
SOURCE Rattus norvegicus (strain:Brown Norway) liver hepatocyte DNA,
clone:773h04.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (sites)
Watanabe,T.K., Hishigaki,H., Okuno,S., Mizoguchi,A., Oga,K.,
Tsujii,A., Ono,T., Yamasaki,Y., Kanemoto,N., Takahashi,E., Irie,Y.,
Nakamura,Y., Takagi,Y. and Tanigami,A.
The large-scale mapping of rat microsatellite markers
Unpublished (1998)
2 (bases 1 to 222)
Watanabe,T.K.
Direct Submission
Submitted (11-DEC-1998) to the DDBJ/EMBL/GenBank databases. Takeshi
K Watanabe, Otsuka GKN Research Institute, Otsuka Pharmaceutical
Co., Ltd, 463-10, Kagasuno, Kawauchi-cho, Tokushima, Tokushima
771-0192, Japan (E-mail:watanabe@otsuka.gr.jp, Tel:+81-886-65-2888,
Fax:+81-886-37-1035)
FEATURES
source location/Qualifiers
1..222
/organism="Rattus norvegicus"
/strain="Brown Norway"
/db_xref="taxon:10116"
/cell_type="hepatocyte"
/clone="773h04"
/tissue_type="liver"
/note="773h04F-5'-ACTGAGCCAGCGCTCTGCCA-3',
773h04R-5'-AGGGTTCACAGGAGATGAC-3'"

BASE COUNT 54 a 94 c 25 g 49 t

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x A0048829 ..

Align seq 1/1 to: A0048829 from: 1 to: 222

131 LyserTherPherThiUpProvalleu 139
|||||
9 AAATCTACCTTCACTGAGCCAGTCTC 35

seq_name: gb_sts:G24202

seq_documentation_block: 400 bp DNA STS 31-MAY-1996
LOCUS G24202
DEFINITION human STS WI-13519, sequence tagged site.
ACCESSION G24202
VERSION G24202.1 GI:1344528
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the unigene
collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 400)

AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSS
JOURNAL Mapped STSS
COMMENT Unpublished (1995)
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: TCTCTGGAACCTGAGATGAGTGG
Primer B: GCAGCTTGCCTACTGCCAA
STS size: 150
PCR profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 mM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

FEATURES
source Derived from dbEST (genbank accession R42493).
location/Qualifiers
1..400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="203.9 CR from top of Chr5 linkage group"

STS
primer_bind 20..42
primer_bind 20..169
BASE COUNT 97 a 82 c 96 g 123 t 2 others

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x G24202 ..

Align seq 1/1 to: G24202 from: 1 to: 400

755 LeuThrasPserLeuGlyAsnLeuLys 763
|||||
11 TTAACTGACCTCTCTGGAACCTTGAG 37

seq_name: gb_ov:HEF150

seq_documentation_block: 599 bp mRNA VRT 22-OCT-1995
LOCUS HEF150
DEFINITION Heterodontus francisci tcr beta gene.
ACCESSION L47474
VERSION L47474.1 GI:1032339
KEYWORDS T cell receptor beta chain.
SOURCE Heterodontus francisci (tissue library: lambda zap) adult spleen
cDNA to mRNA.

ORGANISM Heterodontus francisci
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Elasmobranchii; Galeomorphii; Heterodontoidae; Heterodontidae;
Heterodontus.
1 (bases 1 to 599)
REFERENCE Hawke, N.A., Rast, J.P. and Litman, G.W.
AUTHORS Extensive Diversity of Transcribed TCR Beta in a Phylogenetically
TITLE Primitive Vertebrate
JOURNAL Unpublished (1995)
FEATURES Location/Qualifiers
source 1..599
/organism="Heterodontus francisci"
/db_xref="taxon:7792"
/dev_stage="adult"
/tissue_type="spleen"
/tissue_1ib="lambda zap"
309..599
mat_peptide
/partial
/note="incompletely rearranged putative D-J-C from mRNA;
mat_peptide showing constant region"
BASE COUNT 172 a 147 c 149 g 131 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x HEP150 ..
Align seg 1/1 to: HEP150 from: 1 to: 599

936 LeuLYsAsnPheGlnGlnLeuAsnLeu 944
|||||
126 CTTAAAAACTTTCAGCAGCTCATCTA 152

seq_name: gb_pl:AF030304

seq_documentation_block:
LOCUS AF030304 1099 bp mRNA PLN 02-APR-1998
DEFINITION Arabidopsis thaliana zinc finger protein mRNA, partial cds.
ACCESSION AF030304
VERSION AF030304.1 GI:3004881
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1099)
REFERENCE Bloddeau, P., Luo, M., Dennis, E.S., Peacock, W.J. and Chaudhury, A.M.
AUTHORS Cloning of a cDNA encoding a two fingered C2H2 zinc-finger protein
TITLE from Arabidopsis thaliana (Accession No. AF030304) (PCR98-048)
JOURNAL Plant Physiol. 116, 1193 (1998)
REFERENCE 2 (bases 1 to 1099)
Bloddeau, P.
AUTHORS Direct Submission
JOURNAL Submitted (17-OCT-1997) Division of Plant Industry, CSIRO, Clunies
TITLE Ross Street, Canberra, ACT 2601, Australia
FEATURES Location/Qualifiers
source 1..1099
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
<1..889
/note="ATZFI"
/codon_start=3
/product="zinc finger protein"
/protein_id="AAC09174.1"
/db_xref="GI:3004882"
/translation="FVELLFGIFQSKPKPKSCSYRDDDDDDGDEDDDDDDDDVDV

PSSKDIRLEENECENRNVIRPSINSVAPNDQRFGEVDPSSSSLSKDETAALCPFP
HSQYPOSDYQLOLQVOTQOMLPKPSYQCNVGRRELPYQALGKHASRRTPVE
NATGEKMPKILAPSGKIKKSCICHREPSGSLGKHRLHIEGVLRGHRKROEEAV
SOGDKLSPSGNGSVVTHVPDEKQSRKGLIVNKVPPEFNDPGDKDLEGSALLANK
LEQDGIITSTMINGLKFENFL"
BASE COUNT 326 a 214 c 250 g 309 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AF030304/rev ..
Align seg 1/1 to reverse of: AF030304 from: 1 to: 1099

139 LeuTPARGLYsAspGlnHsHsHs 147
|||||
147 CTTTGAGAGAGATCATCATCATCAT 121

seq_name: gb_htg:AC051908

seq_documentation_block:
LOCUS AC051908 1100 bp DNA HTG 15-APR-2000
DEFINITION Giardia intestinalis clone KJ3172 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC051908.1 GI:7575228
VERSION AC051908
KEYWORDS HTG; HTGS-PHASEO.
SOURCE Giardia intestinalis.
ORGANISM Giardia intestinalis
Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
1 (bases 1 to 1100)
REFERENCE Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.O., Kim, U.,
AUTHORS Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.
TITLE Giardia: a model for ancient eukaryotic genome analysis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1100)
Kim, U., Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.O.,
AUTHORS Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.
TITLE Direct Submission
JOURNAL Submitted (15-APR-2000) Josephine Bay Paul Center for Comparative
AUTHORS Molecular Biology and Evolution, Marine Biological Laboratory, 7
MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT * NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 1100: contig of 1100 bp in length.
FEATURES Location/Qualifiers
source 1..1100
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="KJ3172"
BASE COUNT 252 a 241 c 253 g 354 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AC051908

Align seg 1/1 to: AC051908 from: 1 to: 1100

811 11evallysSerLeuserSerGIuPro 819
|||||
914 ATCGTCAATCCGTATCATCAGAACCC 940

seq_name: gb_Pr:AF181722

seq_documentation_block:

LOCUS AF181722 1382 bp mRNA PRI 10-JAN-2000
DEFINITION Homo sapiens R02AS (R02) mRNA, complete cds.
ACCESSION AF181722
VERSION AF181722.1 GI:6684531
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Van Den Eynde,B.J., Gaugler,B., Probst-Keppler,M., Michaux,L.,
Devuyst,O., Lorge,F., Weynants,P. and Boon,T.
A new antigen recognized by cytolytic T lymphocytes on a human
kidney tumor results from reverse strand transcription
J. Exp. Med. 190 (12), 1793-1800 (1999)

JOURNAL
MEDLINE 20069867
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilotte,L.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,
Avenue Hippocrate, 74, Brussels 1200, Belgium

FEATURES
source
Location/Qualifiers
1..1382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p22.1"
1..1382
/gene="R02"
/gene="R02"
738..992
/gene="R02"
/codon_start=1
/product="R02AS"
/protein_id="AAF23613.1"
/db_xref="GI:6684532"
/translation="MDDDAAPRVGGVPAVNHKHLHDGLROVAGPAAAAHLPRMPP
QLASRRERAPLSORPRHQAGSPETNEKLTNPQVKK"

BASE COUNT

355 a 373 c 344 g 310 t

gene
CDS
1..1382
/gene="R02"
738..992
/gene="R02"
/codon_start=1
/product="R02AS"
/protein_id="AAF23613.1"
/db_xref="GI:6684532"
/translation="MDDDAAPRVGGVPAVNHKHLHDGLROVAGPAAAAHLPRMPP
QLASRRERAPLSORPRHQAGSPETNEKLTNPQVKK"

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AF181722

Align seg 1/1 to: AF181722 from: 1 to: 1382

451 AAGIYAGTARGLeuserSerleu 459
|||||
856 GCTGGCCGCCGCTCAGCTCGTCT 882

seq_name: gb_P1:PSCC266

seq_documentation_block:
LOCUS PSCC266 1751 bp mRNA PLN 12-MAR-1992
DEFINITION P. sativum mRNA of cDNA clone 26g.

ACCESSION X54359
VERSION X54359.1 GI:20680
KEYWORDS alpha-subunit; beta subunit; H-NHase gene; nitrile hydratase;
water-deficit.

SOURCE
ORGANISM Pisum sativum
pea.

REFERENCE
AUTHORS Guerrero,F.D., Jones,J.T. and Muller,J.E.
TITLE Turgor-responsive gene transcription and RNA levels increase
rapidly when pea shoots are wilted. Sequence and expression of
three inducible genes
Plant Mol. Biol. 15 (1), 11-26 (1990)

JOURNAL
MEDLINE 91355842
AUTHORS Jones,J.T.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1991) J.T. Jones, Texas A&M University, Dept.
Biochemistry and Biophysics, Biochemistry and Biophysics Bldg.,
College Station, TEXAS, 77843-2128, USA
See also X54357 and X54358.

COMMENT
FEATURES
source
Location/Qualifiers
1..1751
/organism="Pisum sativum"
/strain="Progress No.9"
/db_xref="taxon:3888"
/clone="26g"
/tissue_type="shoot, wilt distressed"
/dev_stage="seedling"
87..1613
/note="508 aa peptide"
/codon_start=1
/protein_id="CAA38243.1"
/db_xref="GI:20681"
/translation="MGSDSNNGLEKEIGIGATNIGSFINGOKANGCPVHVSPTN
QVYASTYATLDDYEGELRASSAATWTRVPAPRKGEIYRQIGDLRAKLDPLGLY
ALEMGKILABGIGEOVEIIDMCYVGLSRQLGSIIPSEPRHMEVWNPGLGVG
ITAFNPPCAVLGNMNAICVAGTVMVKGAPPTPLITVAVTKLIAEVEFRNNIPGAI
TALCGGADIGHAIKADTRIPLVSTGSSKVGALVOQVOREGKTLERGNALIVM
DDADITLAVRSIFFAVAGTACORCTCRRLYLHESYAVANLEQTLATYKOVKNP
EGTIVGLPHTRSVAVENKNGISAIKSGGKIYVGSVLESEGFVPTIYVTSADAV
VKEELPAPVLYVKFKDLDEALTLNNSVQGLSSIFTRKPSITFRWIGPSSDCIV
NNIPTNGAIEIGAFGEKATGCGREAGSDSWKQYMRSTCTINIGSELPLAOGINFG"

CDS

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x PSCC266

Align seg 1/1 to: PSCC266 from: 1 to: 1751

876 AsnValLeuGluGlnLeuThrAlaLeu 884
|||||
1029 AACGCTTGAAACACTTACCGCACTC 1055

seq_name: gb_Pr:AK027036

seq_documentation_block:
LOCUS AK027036 2003 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23383 fis. clone HEP16466, highly similar to
AF181721 Homo sapiens R02S mRNA.
ACCESSION AK027036

```

VERSION      AK027036.1 GI:10440050
KEYWORDS     oligo capping; fts (full insert sequence).
SOURCE       Homo sapiens hepatoma cell_line:HepG2 CDNA to mRNA, clone_lib:HEP
              clone:HEP16466.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (sites)
              Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
              Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
              Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
              NEDO human cDNA sequencing project
TITLE        Unpublished (2000)
JOURNAL      2 (bases 1 to 2003)
REFERENCE    Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
AUTHORS      Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE        Direct Submission
JOURNAL      Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
              Sugano, Institute of Medical Science, University of Tokyo,
              Laboratory of Genome Structure Analysis, Human Genome Center,
              Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan
              (E-mail:cdna@eims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
              Fax:81-3-5449-5416)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
              International Trade and Industry of Japan; CDNA full insert
              sequencing; Research Association for Biotechnology; CDNA library
              construction, 5'- & 3'-end one pass sequencing; Department of
              Vitrology and Human Genome Center, Institute of Medical Science,
              University of Tokyo (partly supported by Science and Technology
              Agency).
FEATURES     source
              1..2003
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /cell_type="HepG2"
              /cell_line="hepatoma"
              /clone="HEP16466"
              /clone_lib="HEP"
              /note="Cloning vector pME18SFL3"
              misc_feature 1..2003
              /note="highly similar to AF181721 Homo sapiens RU2S mRNA"
BASE COUNT   662 a 388 c 525 g 428 t
ORIGIN
alignment_scores:
      Quality: 9.00      Length: 9
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AK027036/rev ..
Align seg 1/1 to reverse of: AK027036 from: 1 to: 2003
451 AlAGlYArGArGlEuSerSerLeuLeu 459
|||||
111 GCTGCCGCCGCTCAGCTCGCTT 85
seq_name: gb_pr:AF181721
seq_documentation_block:
LOCUS      AF181721 2167 bp mRNA PRI 10-JAN-2000
DEFINITION Homo sapiens RU2S (RU2) mRNA, complete cds.
ACCESSION  AF181721
VERSION    AF181721.1 GI:6684529
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 2167)
AUTHORS     Van den Eynde,B.J., Gaugler,B., Probst-Keppler,M., Michaux,L.,

```

```

TITLE        Devuyst,O., Lorge,F., Weynants,P. and Boon,T.
JOURNAL      A new antigen recognized by cytolytic T lymphocytes on a human
MEDLINE      kidney tumor results from reverse strand transcription
AUTHORS      J. Exp. Med. 190 (12), 1793-1800 (1999)
REFERENCE    20069887
              2 (bases 1 to 2167)
REFERENCE    Van den Eynde,B.J., Gaugler,B. and Pilotte,L.
AUTHORS      Direct Submission
JOURNAL      Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,
              Avenue Hippocrate, 74, Brussels 1200, Belgium
FEATURES     source
              1..2167
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="6"
              /map="6p22.1"
              1..2167
              /gene="RU2"
              303..1733
              /gene="RU2"
              /codon_start=1
              /product="RU2S"
              /protein_id="AAF2612.1"
              /translation="MSGSSARSLSHPVKSVLVYRNGDPPFVAGRRVYHEKKVSSP
              EVFLKEVGVQAEFGAVRNITPTGRIRKLDIOSGNYVAGCGAEFKLVLDI
              GEIKRPMEEVNTVEKPVHSIRVNSAFRRKLDPECTFLIANGDILINPARRLLIPR
              KTIQMDHVLQMTKEITLRSQAVHRLTYLLEKSGAELENGOAFYAVARDKFKKI
              PYGELDPKSTMRPRPGFKASSLPPIVSRKSKSGGNDNRHRSKSYGSDNNSPPPLKR
              KKKEDVSEKLIKQNVKLNKSOETIPNSDEGIFRGAERSETRGAAYOEDEDTO
              VEVPDQAPAEIVDEEDGERANKDAEQKEDPSGMNGDLEEGREATDAEYVEEIL
              DHSQDAPARVNGTDENGEELQVNNELQVLVDKERKSGQSGSGDEADVPDPRP
              PRPEKITSPEENENQNKDYAAYA"
BASE COUNT   675 a 440 c 576 g 476 t
ORIGIN
alignment_scores:
      Quality: 9.00      Length: 9
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AF181721/rev ..
Align seg 1/1 to reverse of: AF181721 from: 1 to: 2167
451 AlAGlYArGArGlEuSerSerLeuLeu 459
|||||
295 GCTGCCGCCGCTCAGCTCGCTT 269
seq_name: gb_pr:AK027029
seq_documentation_block:
LOCUS      AK027029 2290 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23376 fts, clone HEP16231, highly similar to
              AF220185 Homo sapiens uncharacterized hypothalamus protein HT011
              mRNA.
ACCESSION  AK027029
VERSION    AK027029.1 GI:10440041
KEYWORDS   oligo capping; fts (full insert sequence).
SOURCE     Homo sapiens hepatoma cell_line:HepG2 CDNA to mRNA, clone_lib:HEP
              clone:HEP16231.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (sites)
              Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
              Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
              Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
              NEDO human cDNA sequencing project
TITLE        Unpublished (2000)
JOURNAL      2 (bases 1 to 2290)

```


AUTHORS Sugano, S., Suzuki, Y., Ota, T., Ohayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T., and Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Submitted

COMMENT

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'-83'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

Location/qualifiers

```

source
1. .2290
/organism="Homo sapiens"
/db_xref=taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP16231"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
1. .2290
misc_feature
/note="highly similar to AF220185 Homo sapiens
uncharacterized hypothalamus protein HT011 mRNA
BASE COUNT 690 a 503 c 477 g 620 t
ORIGIN

```

```
alignment_scores:
  Quality: 9.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Length: 9
  Gaps: 0
  Percent Identity: 100.0000
```

alignment_block:

US-09-697-089-2 x AK027029/rev .

Align seg 1/1 to reverse of: AK027029 from: 1 to: 2290

```

700 ValAlaGlySerLeuSerLeuValLeu 708
      |||||||
211 GTTGGTGGAGCCTTCTCTGCTTTTG 2085

```

seq_name: gb_ro:RATSVPIIA

seq_documentation_block:

LOCUS	DEFINITION	27-MPR-1993
DNA	4161 bp	
RATSVPIIA		
Rat seminal vesicle secretion	II protein (SVS II) gene, complete	
ACCESSION	J053443 M54795	
VERSION	J053443.1 GI:207114	
KEYWORDS	seminal vesicle secretion	II protein.
SOURCE	Rat (strain CHARLES RIVER)	male seminal vesicle epithelial cell

REFERENCE 1 (bases 1 to 4161)

AUTHORS

TITLE	ABSTRACT	KEYWORDS
Structural characterization of the rat seminal vesicle secretion II: protein and gene		

J. Biol. Chem. 265, 9896-9903 (1990)
MEDLINE 90277684

COMMENT

by S.E.Harris, 25-APR-1990.

FEATURES	Location/Qualifiers
source	1. 4161
	/organism="Rattus norvegicus"

	/db_xref="taxon:10116"
CAAT_signal	1934 . 1946
TATA_signal	2013 . 2020
exon	2043 . 2137

```

/note="seminal vesicle secretion II protein (SVS II)
precursor"
/number=1
join(2065, .2137,2377, .3548)

```

```

/note="seminal vesicle secretion II protein (SVS II)
precursor"
/codon_start=1
/protein_id="AA42192.1"
/db_xref="GI:207115"
/translation="MKSSVFILSLFLLEKRAAVYGVYGTGKHFQSSSGFMGQK
HLNLFELSGSEAEAEISLFMSQHMFQDGDGMAQTSVSEHTGVKCAALCRGVYS
OLKSQESOLIKSFQRYKSSGGLKSGGSLKSGVYKSSSOLKSFQYVAASSQLK
QKAVAGSGLKLYGDKMSGQYKSFQKSSGSGVYSFGOMKASQSIKTSFGQRSCQ
GOLSYGQKMSYVGOTKSLSESKAKSFQYKSGVYKSSGSGVYKSGVYKSGVYK
GLKSYGQKMSQSSSFQYKSGVYKSGVYKSGVYKSGVYKSGVYKSGVYKSGVYK
QPDDEDSVQKQSTQMKTEBDSIFQGGQRQFGQERSYKGYLQYKKKIQEEDQKQ
NENQDPTTKGAGGLIQQLKG"
2065..2130
sig_peptide

```

```
alignment_scores:      quality: 9.00
                       ratio: 1.000
Percent Similarity: 100.000
                       gaps: 0
Percent Identity: 100.000
```

alignment_block:

US-09-697-089-2 x RATSVP11A .

Align seg 1/1 to: RATSVP11A from: 1 to: 4161

1006 GlnPheaspaspaspLeuSerVal 1014
 |||||
 3306 CAATTGACGATGATGACCTCTCTGTA 3332

seq_name: gb_pr:AF181720

seq_documentation_block:

LOCUS	AF181720	4377 bp	DNA	10-JAN-2000
DEFINITION	Homo sapiens RUZAS (RUZ) gene, complete cds; and RUZS (RUZ) gene, partial cds.			
ACCESSION	AF181720			
VERSION	AF181720.1	GI:6684526		
KEYWORDS	human.			
SOURCE				

AUTHORS Van Den Eynde, B. J.,

kidney tumor results from reverse strand transcription
J. Exp. Med. 190 (12), 1793-1800 (1999)

MEDLINE	20069887
REFERENCE	2 (bases 1 to 4377)

AUTHORS

TITLE	Direct Submission

Avenue Hippocrate, 74, Brussels 1200, Belgium

FEATURES

Location/Qualifiers

source 1. .4377
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p22.1"
1754. >2348
/gene="R02"
/product="R02S"
1754. .>2348
/gene="R02"
complement(<1912. .>2166)
/product="R02S"
complement(1912. .2166)
CDS
/codon_start=1
/product="R02AS"
/protein_id="AA023611.1"
/db_xref="GI:6684528"
/translation="MDDDAAPRVGCPVAVHKHNLHDLGQVAGPGAAAHLPRWPP
QLAASREARPPLSORPHRTGAGSPETNEKLTNPQYKER"
2056. .>2348
/gene="R02"
/codon_start=1
/product="R02S"
/protein_id="AA023610.1"
/db_xref="GI:6684527"
/translation="MSGSSARSSHLSPVVKSVLYRNDPFYAGRRVTHKKVSSF
EVFLKEVTGGVQAPFGVARNIYPTGRHRIKLDQIQSGNIVAGGQEAERKLT"
BASE COUNT 1296 a 858 c 828 g 1380 t 15 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AF181720/rev ..
Align seg 1/1 to reverse of: AF181720 from: 1 to: 4377

451 AAGAGTARGATGleuSerleuLeu 459
|||||
2048 GGTGGCGCGCTCAGCTGCTGCTT 2022

seq_name: gb_pr:AF106066

seq_documentation_block:
LOCUS AF106066 9968 bp DNA PRI 04-JUL-1999
DEFINITION Homo sapiens RAD17 pseudogene, complete sequence.
ACCESSION AF106066
VERSION AF106066.1 GI:5353548
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 9968)
Bao,S., Sun,Y., Chang,M.-S., Wong,W.-K., Zhang,Z., Liu,Y., Li,J.,
Wang,Y., Xiao,L., Gravelin,L., Sutherland,R., Weisberg,E.,
Cheng,E.Y.S., Lo,K.-M., Aulclair,D., Hao,L., Sasaki,H., Sonoda,H.,
Campbell,M.S., Kraeft,S., Reddy,E.P. and Chen,L.B.
Hrad17, a human homologue of the Schizosaccharomyces pombe
checkpoint gene rad17, is overexpressed in colon carcinoma
Cancer Res. 59 (9), 2023-2028 (1999)

TITLE
JOURNAL 99247539
MEDLINE 10232579
PUBMED 2 (bases 1 to 9968)
REFERENCE Sun,Y.P., Wong,W.-K. and Liu,Y.
AUTHORS Direct Submission
TITLE Submitted (13-NOV-1998) Cancer Biology, Dana-Farber Cancer
JOURNAL Institute, 44 Binney Street, SM 1058, Boston, MA 02115, USA
FEATURES Location/Qualifiers

source 1. .9968
/organism="Homo sapiens"
/db_xref="taxon:9606"
7622. .9826
/gene="RAD17"
/pseudo
BASE COUNT 3477 a 1765 c 1714 g 3012 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AF106066 ..
Align seg 1/1 to: AF106066 from: 1 to: 9968

517 CysLeuGlyLeuSerIleAlaIys 525
|||||
7048 TGTCTGCTAGGCTTGACATAGCAAA 7074

seq_name: gb_ba:AE000560

seq_documentation_block:
LOCUS AE000560 10351 bp DNA BCT 06-APR-1999
DEFINITION Helicobacter pylori 26695 section 38 of 134 of the complete genome.
ACCESSION AE000560 AE000511
VERSION AE000560.1 GI:2313554
KEYWORDS
SOURCE Helicobacter pylori 26695.
ORGANISM Helicobacter pylori 26695.
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
Helicobacter.
REFERENCE
AUTHORS 1 (bases 1 to 10351)
Tomb,J.-F., White,O., Kurlavage,A.R., Clayton,R.A., Sutton,G.G.,
Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
Dougerthy,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D.,
Hickey,E.K., Berg,D.E., Gocayne,J.D., Uitterback,T.R.,
Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
Venter,J.C.
The complete genome sequence of the gastric pathogen Helicobacter
pylori
Nature 388 (6642), 539-547 (1997)
JOURNAL 97394467
MEDLINE Erratum: [[published erratum appears in Nature 1997 Sep
REMARK 25,389(6649):412]]
REFERENCE 2 (bases 1 to 10351)
Tomb,J.-F., White,O., Kurlavage,A.R., Clayton,R.A., Sutton,G.G.,
Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
Dougerthy,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D.,
Hickey,E.K., Berg,D.E., Gocayne,J.D., Uitterback,T.R.,
Peterson,J.D., Kelley,J.M., Cotton,M.D., Meldman,J.M., Fujii,C.,
Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
Direct Submission
Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 10351)
REFERENCE White,O.
AUTHORS Direct Submission
TITLE Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
1. 10351
/organism="Helicobacter pylori 26695"
/strain="26695"

```
gene
/ob_xref="taxon:85962"
/complement(88..273)
/gene="HP0445"
CDS
/complement(88..273)
/gene="HP0445"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0445"
/protein_id="AAd07521.1"
/ob_xref="GI:2313562"
/transl_table="MOFLPKDLRISQKYVVPKGFDPNKEYKFLVKDLPKLHREV ELYMKLFEINEVDRD"
/complement(292..1071)
/gene="HP0446"
/complement(292..1071)
/gene="HP0446"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0446"
/protein_id="AAd07516.1"
/ob_xref="GI:2313557"
/transl_table="MOERYKRVLDANILKEHANNVCYSKHSDRFIPKDFEGY VGCLEKTLINPSTNTPFGCGCKMELLSFIVDSRSKIALNRHATPLQDC RTRRNHQLERIGFTAFKNNKGYIFLSMHDPTKIENTEHSGLNCSNLPNSFPGI CYGLKLFKRLKDRARLANKGVEFLKIPKHPQWQNFNNILEFVSVNPPCSYRLT YDQLVGKKIPNKISYQDAIVKKNHRADELSDKEILAS"
1179..2261
/gene="HP0447"
1179..2261
/gene="HP0447"
/ob_xref="GI:2313558"
/note="similar to SP:P34243 PID:395256 PID:486007 percent identity: 38.22; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAd07514.1"
/ob_xref="GI:2313555"
/transl_table="MSLANDLEILSSOSPNDKRIHVLLEYISVLETPMDLESLRD YNFVNTTGOHNOALERKETPFDSVIVDEAKANPLELIMMALAKRITLIVGDR OUPHYLDDIEIKKLEDESODADEIKALDSMERKIKERAKOTKELDEKCEITLNM QGMRNPLGELVSDVYKPKHNSFESPLGKHLEKPFKNHNEVDNKNLAMDVAVES KERNADGYSYRSEETAIKEDLCEIDLEPKDEPDTFGVITFFSEOKRLLEQALGVANL EIGTVDSFGKEFDVFLSSVTRHTEGFGFLKISSCLVALSRORALIVAGDREKF DPESKDKVSGLEFELQSCKEGKITL"
2258..2644
/gene="HP0448"
2258..2644
/gene="HP0448"
/ob_xref="GI:2313558"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0448"
/protein_id="AAd07517.1"
/ob_xref="GI:2313558"
/transl_table="MKIISFENDSVCDKHLILPCRIYCELEIRHNDLUNALKECAL KLESVGVDEELKGLFGEGELDLLEPLEIKETITKLEEKIKIHAFYVNLINQERFLH FVDLEVRRAIDCKEVKCAPAKDWMAS"
2891..4309
/gene="HP0449"
2891..4309
/gene="HP0449"
/ob_xref="GI:2313558"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0449"
/protein_id="AAd07518.1"

gene
/ob_xref="GI:2313559"
/transl_table="MEFPADSSPKAEASTLYPKKNNLSVNEIIRKTIETSIKYYHO DILRHAKVSMHVGCVYLLHAKAPDSSKIEVNECNHPIEKMAELKETNPENPEPK REKOTVOTESKPSRKIKVFNKNTAPDDLYITECNKLDTHQKFEITGALDYIE QIFDETPSLKGDLETYKKAODLSKRPMLGFQDISDLNDSRLNALCYEKVQALA KYPSFLQDLHDIKKRYKNDKNGEKPDGSSLTRVLEERYRGDIYFLVNLNLPILK ERENAOEKHYKKNAEIDDRSOLSNNAKPLFECEGVNALAKAYSKKEEKEKML ILNFEKIFETILFEAIOKOPKDPKDELDLDYPLKMLDSIREVREDPLRAKEND EASIGAYVVLSSCKEYESVEKVDWDFIARLLALRQHVHKITKELERLEEDLEK LERQALEFNKIANKIYLEKR"
4306..4440
/gene="HP0450"
4306..4440
/gene="HP0450"
/ob_xref="GI:2313563"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0450"
/protein_id="AAd07522.1"
/ob_xref="GI:2313563"
/transl_table="MSGNELELRARETELKRVAVELDKREKQKIDDAFQOEKFI "
4451..4835
/gene="HP0451"
4451..4835
/gene="HP0451"
/ob_xref="GI:2313563"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0451"
/protein_id="AAd24378.1"
/ob_xref="GI:4580426"
/transl_table="MILETISKQLOEQESLNNKDYRKLESIKERSDRLESQTOFD AQLEKWEONKGEQOITDILNREQDLDIDQORLNKOKALENHMKOREE"
4911..6467
/gene="HP0452"
4911..6467
/gene="HP0452"
/ob_xref="GI:2313556"
/note="similar to GP:2983256 percent identity: 34.34; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAd07515.1"
/ob_xref="GI:2313556"
/transl_table="MADETNALRQKNRELNKRIDMQKDYDREKLERNDLEQCKEDL LLAENLNRRIQEMENENKLDPRDERIKLEBEKRELGLIAOKENAROKYNTLSV KNOLEAELMLNMEKPEKLNMYAGVEDPEKROKNTKEOIVKTNPVYLGAPSNVEVLA FLERIKGMOERNVYPRKULIMHRTALSTSLSPSLVSGVSGTGSSELPRLYHFG GLNFIATAOPTWDSPESLMGTFHALENKFDATFELRFTIQTLSNNEPVGLEKAMS IYLDENMLAHLELFAEFLSKLEIKCSQETNISIKLGTGLWEPLGDNLLFEGTIN EDESTKMLNDKYLDRAFCLNFRPKITLKGQOKPIPSNDGYLVEFTFNKIMKQOE AKLEGGYKLTIEINERLIDAGRSIGHRVWQMSPTVMHNPVLVLAHSDKNALQAYE ECLVLIKPEKRLGVQTRNNQHLTKIDDLKDPSSVSDFFQAMENDEKKEVPFSTNVLN NYEYKLTLLK"
6474..9539
/gene="HP0453"
6474..9539
/gene="HP0453"
/ob_xref="GI:2313560"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP0453"
/protein_id="AAd07519.1"
/ob_xref="GI:2313560"
/transl_table="MDLELYAVPNHIERLKARSPLRSTAFPDQFASPEYRDLFSYLE NIVQDPYEKKPKKDDLYFLCKEYVEPRALKAFESNLNTNITRKHLKPLEKARFQKCA LDIAKPRGSLKELCDNKSIVSVKRVNNTNHEENPLKFEIKELRTIHWREITEQOV FBEILFISITSPNGVAQOIDEKQALIPNNLHAPKHXVRIKRAHMDWLVDGYSIMNL DQIFYLECLYQAOQFYTSKNIIEPTLLIRNEODLYALIKNSPPIKOLDSFEKMRKAKFEFE
```

NELRPIULDEIPELELCCKYKEMIDMSPEFPALLVNGNEKILKPLVKKO
ENNTIINANGAGKIDEGYLANALKNTDEIVLVEFMDFEKERYIEKLYYLDNLIK
NEEPARIKHAKISLYEKDAKFPKSVYALGESSLEFNKRLKNELRNYSVDLVYKHNKS
KTFNDGCVLEKROKSDSKREPLILODSFIKALKNFKRALGKEGFLIKCPLKLS
MEVVDGFRKNEIKDKTIIGDKTEIEFPFIIPKGRESFALPLINEKELAVGK
ITSKDFLENDDEEYKTLTYDIGTEFNVLEKFPVNDLKPIYMEMORIDRELEPTP
SIKRSIDELKNDKPRKRSKSDLEFMALEBQETIKDINSPPREYLEKKEGGISII
GEDRNNEIYIMETNGKRVFCHSRCKGKSVNDELISLARVCLEVGPKNDHGRKRG

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AE000560 ..

Align seq 1/1 to: AE000560 from: 1 to: 10351

735 ValThrAsnLeuYstHrLeuSerIle 743
|||||
9108 GTTACTAACCTAAAGACTTTAGTATT 9134

seq_name: gb_ba:AE004624

seq_documentation_block:
LOCUS AE004624 11533 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 185 of 529 of the complete
genome.
ACCESSION AE004624 AE004091
VERSION AE004624.1 GI:9947973
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa.
Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 11533)
Stover,C.K., Phan,X.-Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Huinagle,W.O., Kowalik,D.J., Lagrou,M.,
Gardier,R.L., Goltz,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Westbrock-Wadman,S., Lam,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
2 (bases 1 to 11533)
Stover,C.K., Phan,X.-Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Huinagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltz,L., Brody,L.L., Coulter,S.N.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Tolentino,E.,
Folger,K.R., Kas,A., Larbig,K., Lam,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sailer,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
LOCATION/Qualifiers
1. 11533
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
164. 1363
/gene="PA1974"
164. 1363
/gene="PA1974"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05362.1"
/db_xref="GI:9947974"
/translation="MNKNNTAVRRLPLTLASLGLGIPRAADILLYXDETTTST

gene
CDS
1466. 2629
/gene="PA1975"
1466. 2629
/gene="PA1975"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05363.1"
/db_xref="GI:9947975"
/translation="MEGOGEGVCGVCGSSARDVEYVADLAROLIHPLHVLFFCSA
EYDPLALAMELERYTGGIDLYGCTTAGETITAGYGRGVSVAGDEYRSFALISALIDE
MERESLDAAQGMVEFTLVAGCRGGIAPIKDHSFALTLDLSSREVVLAALGALGA
.1PHFGSGAGDDRHILHTHTVHQGFHTGAQVAVLVNTWLDREVEFTTHAVRAKLVV
TRADSGSRVHELNALPALEAYARLVGAPOALDPALESAHPLAVRVDERYYSIQ
ANEDLSLFFCAVENGIYLTAMSTGPIPLNLEAOPRRLHERLHPLLTIGDCFLRL
ELEVGCDTQATAAFILRSQVYIGFNTYGGFNGMHNITFTGVALGRPGVRR"
2559. 5204
/gene="PA1976"
2559. 5204
/gene="PA1976"
/codon_start=1
/transl_table=11
/product="probable two-component sensor"
/protein_id="AAG05364.1"
/db_xref="GI:9947976"
/translation="MACISTRPSDPSLPADPAEYSAELQARVVALLEERHKLQRIQ
ALIGRESSAOSDEAGAFQYSVYLAOVERDPALEQTMLEKASQULSDRLRA
ETAOHILDAIESISDAVLPDPOORIVLFNERFRALIDSRRARIGPMRARELRLA
RSGGLIVREGGDOQTYRLRDERMLMSRPRDGLAVIYSDYETIKSEIARE
QALAKSHLORAVDNLSQVAMGADVLELMNRFLLEGLAPIDAGHAKFAEMAE
SELPLITRSDSGSPISAREQRLPDERVLEIRHALPTGFVNTFDITERVHAE
ALRKSERIRLITDPOPALIAYISDELVEETNKYERMYPMPSGSMGRCLRYHGA
EHCRLPEYFEPALGESVTEIEATLSEGERVWORSVYRNQADGVAGLPIVIR
ITERRTAEALHOAVONLEORRETRTALTLINDQKEIHERSOVEARLEAREAE
TANISKTEFLAASHDILQPLNARLFTSALOERGSATRONDEALVNNISLSEVEN
LGTIVDISKLAVGVTTPDIAPFELGELLGNLAERYOYVASEGLRDPVACGALYRS
QOLARILRLNLSNARTYTPRGVYLGRKRRQRUSTEVMDTGVIADKIGETFOER
KRGESORCHODRGGLGAIYDKIARMGHVRVYSGKSCFAIEVPLARHPRSR
AEPASADLDERLERLGRSVWVLDNAACACMRITLLEAMGCRVYTALEEDLAVOVN
YHAEADLIVYHLDORNGVDVAVAIINARGSPPLPALMITANYSNELKQVRELHGT
LMHKVPRMKIKTKALCHLREASV"
complement(5206. 6069)
/gene="PA1977"
complement(5206. 6069)
/gene="PA1977"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05365.1"
/db_xref="GI:9947977"
/translation="MPDRRTLLTVLAMLAFAGNSILCRAALKDPAIDAVSFALRLF
SGAMLAVLHLRRPATPARGWRGGAALLVYVAASPYAVVODAGGALLGAVG
VTLIAGLIGRGERGCGALIGFLTAGLLPLLPASAPPLGAILMLLSGLIAGLY
TLIRGGGDPLAVSAGNPLRALAFALLLFLFHQGLRDGGLAVALLSGLASGLG
AWYSALPGLTAIQASVQSVPIYALCGALLGELPAPRLPPTTAVLIGIALIITA
PRLGARAREGA"
complement(6079. 6744)
/gene="PA1978"
complement(6079. 6744)
/gene="PA1978"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG05366.1"
/db_xref="GI:9947978"
/translation="YKTLIADHPDLFREAIHNVYADGFPGESEWETADLDASGLTIO

```

HHDDLILLLDLMPCMGHGLNGLMLRNEAPITPVIVSAEDQKQVLAITYGAVGF
ITKSSPRAQMTAEIIOILNGNVLPSDVIKRTOKSSPRSGHSEHSGISPELLOLTKRO
LTVLERMTGESKQJALYNLDIAETTVKXHVSAILKLKVHNHVQAILLSAGDIDFAAY
LR"
gene
7409..8059
/ gene="PA1979"
8059..8059
/ gene="PA1979"
/ codon_start=1
/ transl_table=11
/ product="probable two-component sensor"
/ protein_id="AAG05367.1"
/ db_xref="GI:9947979"
/ translation="MDRLPAGPAEYRWMLAVERERRLLAQAHHDDIGOYLIGIRTO
LLIRCTAGDGPALRLRFRYCOGLDSTRALLRGIPRALROSIDELRACVADMO
EEOGJLCELRIGLPLPDLAAGTHLIRLLOELNNVNRHAGASRRIRLQCRGGLY
LSVRDGRCCADPAFGILDSMRORARLGAELSLRSPGGALRLRMLPERD"
8072..8749
/ gene="PA1980"
8072..8749
/ gene="PA1980"
/ codon_start=1
/ transl_table=11
/ product="probable two-component response regulator"
/ protein_id="AAG05368.1"
/ db_xref="GI:9947980"
/ translation="MKVLLYDDHFVYREGSLALRLGLPDVEYENAGDEEALQAVOR
EIPSLVLDLPGISLELTRRLRLDQLRVLFSLDELAIVLQAQADAGRGVIT
KRAAPVLLLEAIRVLAQGLLEQPLATRLAOCQSWEEQVLAALRGITREFEFRLA
RGLAIREVNEHLSAKTVSNHVSILKRLQVSTQAEVLHRAIDSGVLRLGLPLAE
EARGSLG"
complement(8776..9423)
/ gene="PA1981"
complement(8776..9423)
/ gene="PA1981"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAG05369.1"
/ db_xref="GI:9947981"
/ translation="MKTFLCAALLGAGLQAADIDVYIAGCRLEPGARCANADL
RGADRLNDLGRIDLADLSGADLRHARLDLANLEKAMLRGADLRASLQOANLRGA
DLGARAVALIQAMGLFAQAGMAGADLTAAYLEFALISGRGHQATLRAADLEMWLS
RADLGADLRDANLQEVKLAEANLEPADLRGKVRGNFGNSMQCKCKCPAGMW"
complement(9477..11348)
/ gene="exaa"
/ note="PA1982"
complement(9477..11348)
/ gene="exaa"
/ codon_start=1
/ transl_table=11
/ product="guinoprotein alcohol dehydrogenase"
/ protein_id="AAG05370.1"
/ db_xref="GI:9947982"
/ translation="MTTTRSPAPAGILRPSLHCLFAVALSGSAGALADVYMEDIAN
DKRTGDVLYOGMTGTHAORMSPKQVNAADNVEKFLPAMVSGSDEKORQEOAOAISD
GVIVYASYSRLFDALDAKGRKLTMTNHRLLPDDIRCCQNVNRGAIIYDCKPFGID
ASVVALNKNTGYVKKKRRADHAGCTTGATPYTKDCTGVLLIHSGSGEPGYVC
RLFARDPDTGEIIMKRPVEGHMGRUNGKSTVIGVKA.PSWDDNSSTGVEESH
GGGAPQWASAFDAENNTIIVGAGNDPQPMNTWARTAKGNPHDYDSLXTGGVGVPS
GCVKVEYQHTPNDAMDNSGNELVLEFDYKAKDGKIVKATAHADRNCFYVVRSNKL
MAAFPEVDNITWASHIDIKTGRPREREGORPPLPERGOKHGAKEVSPPELGKNNP
NAYSODTGLFVYPAHMKEDVWYMEVSYTKGSAYLGMCFRIRKMYDDHVSIRAMPV
SKRYVMEKEHLPLMAGVLAATGNLVFTGTGCTGRTKAPDAKSGKELMKQTSGLIYSP
PTWEDDGGQYIGVYVGGAVPLWCGDMADLTPVAQSGSFWEKLPMSWNRTRSR"
BASE COUNT 1798 a 3825 c 4087 g 1823 t
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-697-089-2 x AE004624/rev ..
Align seg 1/1 to reverse of: AE004624 from: 1 to: 11533
684 PheserSeraIatnSerIeuArgIeu 692
|||||
1278 TTCAGCTCGCGCACCAGCTTAGGTTG 1252
seq_name: gb_pl:AC074110
seq_documentation_block:
LOCUS AC074110 20509 bp DNA PLN 30-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC T18C15 genomic sequence,
complete sequence.
ACCESSION AC074110
VERSION AC074110.5 GI:11072070
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 20509)
Lin.X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Uteback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nielsen,W.C., and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T18C15 genomic sequence
Unpublished
2 (bases 1 to 20509)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (14-JUL-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
3 (bases 1 to 20509)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (02-NOV-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
4 (bases 1 to 20509)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (30-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
On Nov 2, 2000 this sequence version replaced gi:11024897.
Address all correspondence to: atetigr.org
COMMENT
BAC clone T18C15 is from Arabidopsis thaliana chromosome 1
clone.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan (Chris Burge,
http://CCR-081.mil.edu/GENSCAN.html), GenemarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/Genemark/), Glimmer4 (a variant
of Glimmer, see Mihela Pertea,
http://www.tigr.org/softab/glimmer4.htm/glimmer4.html), and
GeneSplicer (Mihela Pertea and Steven Salzberg, contact
mperle@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arjan Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
source
1..20509

```


alignment_block:

US-09-697-089-2 x AC074110/rev ..

Align seg 1/1 to reverse of: AC074110 from: 1 to: 20509

139 LeuTPPrGlyAspGlnHisHisHis 147

14072 CTTTGAGAGATGATCAATCATCAT 14046

seq_name: gb_in:AF078157

seq_documentation_block:

LOCUS AF078157 38358 bp DNA INV 11-JUL-2001

DEFINITION Caenorhabditis elegans cosmid F25E5, complete sequence.

ACCESSION AF078157

AF078157.1 GI:3335236

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 38358)

The C. elegans Sequencing Consortium.

Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium

Science 282 (5396), 2012-2018 (1998)

99069613

2 (bases 1 to 38358)

Wilson, R. and Bradshaw, H.

The sequence of C. elegans cosmid F25E5

Unpublished

3 (bases 1 to 38358)

Waterston, R.

Direct Submission

4 (bases 1 to 38358)

Waterston, R.

Direct Submission

Submitted (18-OCT-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

5 (bases 1 to 38358)

Waterston, R.

Direct Submission

Submitted (11-JUL-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RO, England

email: tw@nemalode.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.

It may be shorter because we only sequence overlapping sections

once, or longer because we provide a small overlap between

neighboring submissions.

This sequence was finished as follows unless otherwise noted: all

regions were double stranded, sequenced with an alternate chemistry

or covered by high quality data (i.e., phased quality >= 30); an

attempt was made to resolve all sequencing problems, such as

compressions and repeats; all regions were covered by sequence from

more than one m13 subclone.

NEIGHBORING COSMID INFORMATION

The 5' clone is F18E3. 200 bp overlap; 3' clone is C18B10, 200 bp

overlap. Actual start of this clone is at base position 1 of

CELf25E5; actual end is at 38358 of CELf25E5.

NOTES:

FEATURES

source

Coding sequences below are predicted from computer analysis, using

the program GeneFinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers

1..38358

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="v"

/clone="F25E5"

836..1798

/gene="F25E5.8"

join(836..1324,1580..1798)

/gene="F25E5.8"

/note="weak similarity to 'POU' transcription factors"

/codon_start=1

/product="Hypothetical protein F25E5.8"

/protein_id="AAG24071.1"

/db_xref="GI:10864358"

/translation="MKPVIVAIIVSEFSFQHPDPFKASRSTGNPFGSGSGSH
KPEHHHKKHHHSHSLYFGKHSKKHLPEVENDGTHSHSHSDVSSGSSSLKAKKHHL
TAHFQDLFIHGHHHHGHHPHSKONSASTARKNSTGECPCSKOECHHSHNIT
VGADEVPVLAIPNEQNLHAHFTKNOSPCISRYVSSSLKELKHKIKRTYSDDT
DPPEKCSVAASRYVTH"

complement(2400..2813)

/gene="F25E5.9"

complement(join(2400..2687,2742..2813))

/gene="F25E5.9"

/product="Hypothetical protein F25E5.9"

/protein_id="AAG24079.1"

/db_xref="GI:10864366"

/translation="MKQILVIVQVSPADGVIVRSVIRHNGAIDNECPSCVFLR
LPSSSFCDLCAICGAKPTSLIECWCAPHSFPLVSLKRRTRRRREKSTITE
MQGSDCCKGGRKLNH"

3216..6377

/gene="F25E5.7"

join(3216..3355,3541..3916,3963..4110,4292..4452,
4497..4656,5291..5907,5958..6377)

/gene="F25E5.7"

/note="weak similarity to coagulation factor IX"

/codon_start=1

/product="Hypothetical protein F25E5.7"

/protein_id="AAG24078.1"

/db_xref="GI:10864365"

/translation="MGRKKKHNAKIGENANVKTTIMFLIDQLDGFDTKLCVCAEK
YMTTHFYRGQFFDEBDLIVDICHSSIPSQCDARTVEKFWNTTIDIEHVEKRN
CMSCPEFIKMDLISRENNYVNGRIYCVLISEVPKANKKQAVLYEYPSIRMLEI
LVEKSEFHNMLLAFTSKFEKELARGLNTADPYLIEDIPIFKLYMDITYHKO
YSAVYAKDILQIAKRFNMLEVTTQCONVILDLNENKSTKNKTKLAESYLDVRE
LNPNTMTPSARPRMARPAREKQPTDAEBEGISELFAEPWNOCHCKSKAPLOR
RLTNGKLEAFENPMVAVAOFIYVOKDGRKRLVAVFGVYISPHILITLVLSYD
GVFGFQHKKNMNTSGTCLEHYPEEYDTRDILMDRSKFMTRKFDLVLRVAVIN
GCPDVNSAKVNLILELKLPLELPSIISNDPOLFDRSSDFSVSGIDAKVLSGL
FKPVNCSVGPSCAEAVDNKOMCAVDSGSGALVYGVYVGNANCANP
ONLTDKRFININHRESICKLIGVCPDERIDIPSSALPPYPPVGVGDGVLAH
PPMRLNGSDPAESIRAEKIDPMKLNLSHHBEGYKTYNITINNKKT"

10316..11606

/gene="F25E5.6"

join(10316..10720,10773..11074,11123..11229,11347..11606)

/gene="F25E5.6"

/note="weak similarity to steroid/chyroid/retinolic nuclear

hormone receptors"

/codon_start=1

/product="Hypothetical protein F25E5.6"

/protein_id="AAG24077.1"

/db_xref="GI:10864364"

/translation="MOCGVGDCFDLTRIAPVRCACRACFEKCVLGNMNMJOIDEK
ESAKNYKILKIKLKNQYCOKSEHSEKQOTINSIDATETTKOKIVDLSYLETLRK
RUSAVNPHWSTIPCLIEPFLLENKICLADKSGRPMGAPLQTPAPKPIPERRNPVG
RPYFPDKTWLTYDLDTLROSLTEAKTFEFQDTRQLIMRVYSLALNNLHSTYFVA
KKFDITLIPDGSILAPNKGAVYAEIVMSMAPLIRCNIDIEVITLKATICLNPAPEL
STHAQKILAKEREQYADALFDHCIRSRTPAQPAELISMVILBROQRMQDHLHLH

```
gene
12170..13778
/gene="F25E5.5"
join(12170..12508,12706..12945,12995..13192,13236..13370,
13417..13547,13592..13778)
/gene="F25E5.5"
/note="coded for by the following C. elegans cDNAs:
yk38d2.5"
/codon_start=1
/product="Hypothetical protein F25E5.5"
/protein_id="AAG24076.1"
/db_xref="GI:10864363"
/translation="MSENESEFGLNVEFELLDDAEQYSEKNEKPEKSKOEIDELCE
FTGGFEKMPVLTDCKLEVAKYIDLMSKCKLSTCSQDYELVNTALKLDCIDIDQNE
TWAYELSKEDDNVTIRIQSFDPATKTYMLIPSOLEDVQMRVHVHYRNSKPIST
VINSQNYEAARVAEASWLRSEVKEIIVEMSKYPTDTSQIESLPCNVRYGSD
INVMMLRRVPEDLNSLKILREEDRKVFIPDLELPHVMNVSSFEYCGRAFTD
EOLIRLAQMSFPCITYTDCINELPKRMVINGGIGSGFRGYFLMSRPEROILLRC
IEARPMDEAFEEERAGCGDFERVCGRQCEQISRIDPESLTLSISDSSTCIYATG
KRTYEGVYTYDGIPIR"
14304..16051
/gene="F25E5.4"
join(14304..14406,14460..15079,15129..15345,15546..16051)
/gene="F25E5.4"
/note="weak similarity to chymotrypsinogen 2"
/codon_start=1
/product="Hypothetical protein F25E5.4"
/protein_id="AAG24075.1"
/db_xref="GI:10864362"
/translation="MKVSLIVLTITLTAVCAGKIDKEKNEILQIKGIRKNYLFSSGS
OREFINDGTRARGDHPMAVSYVYKANTSKRGVFLGPEITLISARHVLTFNSIKYDDG
RRIIGQEVNAGCNHNEELSDQEMHYHDVDFEHKNDSDRDKRNTASVYITVNGCQ
SSPPATLMEELKESALHNKRGVPCISNSPKHNDSDFEVFLNQGKRVLSGAFEP
TNCATLAPFSCAHAVKONQGLSGDGSASRIADNFTMGFGAQNKNKAPETL
EAFELNGYREIEICEVTGICPTSPSPSESTLSPVPGTLEOEFETKPEPTFD
PIITAESGTTPEITENPVSTYESARTTVMETVEPEDEITESSVRPGESEGPM
SSGSPPEEPSPVDPTQYDPDPTVPAIRIGENOEYVHAETFEPMRVYHNCIS
EETPDNIGCGSVYRGNRGVSDEINITYINAM"
16115..17793
/gene="F25E5.3"
join(16115..17319,17365..17793)
/gene="F25E5.3"
/codon_start=1
/product="Hypothetical protein F25E5.3"
/protein_id="AAG24074.1"
/db_xref="GI:10864361"
/translation="MSVWMLLFSLPIVSCYAPYIKEHARLNTENKLOATCGRMT
RYOKMTLDGERSIESPMWAGSVNYIGIISSTVISPRIILLENLQIANDTLKMTLNFT
VIAQESCKDSKDLLPHQINDMFOYDPAKQMDKAYKHVQRIYITDGCDELDTKPMI
IELBDMFEDKAGHACLPKPTISHSDIQEFTVFGLPNETALISTRYAKACEKEHNG
NSVFCGRPLKGNRRLCAGDFGAGVATIDGRNVVLGIYAEAGNTDCDLEPETOQSAEFT
DLASVTERKLCROTIGYCDOLEETTAEPSTSDGDIYGEESGSESGTEGEQGVTTD
ASIEAIVAEFPEKIKLSCSTFAEKDIIHIRELDVYEKLAKK"
complement(117840..19212)
/gene="F25E5.10"
complement(join(17840..18462,18540..18998,19089..19212))
/gene="F25E5.10"
/note="weak similarity to protein C"
/codon_start=1
/product="Hypothetical protein F25E5.10"
/protein_id="AAG24070.1"
/db_xref="GI:10864357"
/translation="MVSCLFSLVAFITLASSSDVDAFKISEENALOKTCGTRKMRAD
SYTRKVLNGITANAGETPWVALYIPDLHSVYTTGTVNSRHIIISDRFLFTNTTD
GILRHNLKAVIERKMECEGNDYLLPDLVRSVAFPLDLSVERSGHEODLVSVRI
LNGCVKSFQFNRAVAVIELKKPVHKGONARPIGSDRITTYGYRFGYGNRGAVKDA
```

```
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
```

```
US-09-697-089-2 x AF078157 ..
Align seg 1/1 to: AF078157 from: 1 to: 38358
896 LeuSer-SerLeuLeuYshLeuGlu 904
36113 TTCTCTAGTTTATTAAACATTGGAA 36139
seq_name: gb_pr:AF037338
seq_documentation_block:
LOCUS AF037338 39108 bp DNA PRI 22-DEC-1998
DEFINITION Homo sapiens cleft lip and palate transmembrane protein 1 (CLP1M1)
gene, complete cds.
ACCESSION AF037338
VERSION AF037338.1 GI:4039013
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 39108)
Yoshiura,K.I., Machida,J., Daack-Hirsch,S., Patil,S.R.,
Ashworth,L.K., Hecht,J.T. and Murray,J.C.
Characterization of a novel gene disrupted by a balanced
chromosomal translocation t(2;19)(q11;q13.3) in a family with
cleft lip and palate
Genomics 54 (2), 231-240 (1998)
JOURNAL MEDLINE
99047529
2 (bases 1 to 39108)
Yoshiura,K., Machida,J., Daack-Hirsch,S., Patil,S.R.,
Ashworth,L.K., Hecht,J.T. and Murray,J.C.
Direct Submission
Submitted (08-DEC-1997) Pediatrics, The University of Iowa, 140
EMRB, Iowa City, IA 52242, USA
FEATURES
Source
1..39108
location/Qualifiers
1..39108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2-q13.3"
/clone="R30367 from bases 1-36144"
/clone_lib="L119NC03 R chromosome 19-specific cosmid
library"
/note="cosmid library constructed at LNU from flow-sorted
chromosomes from hybrid 5H12-B, which carries one human
chromosome 19 as its only human chromosome"
<190..280
/gene="CLP1M1"
/note="1a: possibly alternatively expressed; detected by
5'RACE procedure and may not be authentic exon"
<190..>38920
/gene="CLP1M1"
<594..664
/gene="CLP1M1"
/note="1b: possibly alternatively expressed; detected by
5'RACE procedure and may not be authentic exon"
<773..804
/gene="CLP1M1"
/note="1c: possibly alternatively expressed; detected by
5'RACE procedure and may not be authentic exon"
933..940
/gene="CLP1M1"
/bound_molec="Sp-1"
951..958
/gene="CLP1M1"
/bound_molec="Sp-1"
protein_bind
protein_bind
protein_bind
mRNA
22922..23039,30798..30883,32035..32155,32759..33003,
33660..33753,35975..36165,36430..36525,36818..36953,
37813..37980,38191..>38920)
/gene="CLP1M1"
/note="possibly alternatively expressed exon 1d"
```


exon
/product="cleft lip and palate transmembrane protein 1"
/gene="CLP1M1"
/note="1d"
Join(975..1046,7542..7654,18666..18789,20018..20176,
22922..23039,30798..30883,32035..32155,32759..33003,
33660..33753,35975..36165,36430..36525,36818..36953,
37813..37980,38191..38477)
/gene="CLP1M1"
/note="possibly alternatively spliced product"
/codon_start=1
/product="cleft lip and palate transmembrane protein 1"
/protein_id="AAC97420.1"
/db_xref="GI:4039014"
/translation="MAAQAADGARSAAVAAAGCGSSGCVTSNGSIGRPPAPTORP
PAQAPNMAOVKGVLFRTITTAISNFRRCAPADODGPGARVARSRLPFDTL
NMLHYISEHEHTDFNATSALEWQHDIYGDWTSGENSGCYEHFAEKLDPQSVQ
KSLIHYFTSGFHPDRKALYRLATVMSRMINKRRRQKRLDGTGEA
DEMIKRAEDYGPVEVISHMNPITINIVDHTPMVKGSPPLDQYFAGSDY
PIIYNDWNLOKDYYPINESLASPLRVSCPLSMRQLAAGSTSPMFLDEL
VRSPEODSVKVALLETPYLLATITIVSYHVEEPLAEKNDIOFMSRSLGSL
VASVFGVQSFVYLLITDNETNVOVSFTIGYLDIMKTKYMDVRIDENRAG
IFRPSFEKSTYIESSTKVYDMAFRYLSWILFPLGAYVSLYLEHKWYSWL
SMLYGFLLTFGFTMTPOLFIYKLSVAHLFWRMLTYKALTFIDDLFAFYIKPMV
YRIGLRDQVVFYLYORMIYRVDPTRVNEFGMSGEDPTAAPVAEVEPTAGALLPT
PAPTPTTATREASTSLPTKPTOGASASASEPQEAAPPKPAEDKKD"

BASE COUNT 8808 a 10190 c 10464 g 9646 t

ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AF037338/rev ..

Align seg 1/1 to reverse of: AF037338 from: 1 to: 39108

450 ThrAlaGlyArgLeuSerSerLeu 458
|||||
31613 ACTGCTGGCAGAGGCTCTCCCTCA 31587

seq_name: gb.pl:AC079674

seq_documentation_block:
LOCUS AC079674 48940 bp DNA PLN 19-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC F10F5 genomic sequence,
complete sequence.
ACCESSION AC079674
VERSION AC079674 GI:12323591
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 48940)
Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Uteback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F10F5 genomic sequence
Unpublished
2 (bases 1 to 48940)
Town,C.D. and Kaul,S.
DIRECT SUBMISSION
JOURNAL Submitted (07-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org
3 (bases 1 to 48940)
Town,C.D. and Kaul,S.
DIRECT SUBMISSION
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712

COMMENT
Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280745.
Address all correspondence to:stet@tigr.org

FEATURES
source
BAC clone F10F5 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), Glimmer2 (a variant
of Glimmer, see Mihela Perlea,
http://www.tigr.org/softlab/glimmer2.htm/glimmer2.html, and
GeneSplicer (Mihela Perlea and Steven Salzberg, contact
mperlea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/t51.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatsMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1..48940
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F10F5"
complement(join(<4473..5050,5131..>5404))
/gene="F10F5.4"
complement(4473..5404)
/gene="F10F5.4"
/note="contains Pfam profile: PF00310 Glutamine
amidotransferases Class-II"
complement(join(4473..5050,5131..5404))
/gene="F10F5.4"
/codon_start=1
/product="hypothetical protein; 5404-4473"
/protein_id="ANG51769.1"
/db_xref="GI:12323593"
/translation="MFAERRMVQKKLKIMARDNRPRPEFGDEEDMYLNCNVA
SRPSLSCDGLVCIPIPGWIVNLPSTGEFLRFSRSGDRPLINDPFCVSKPEVPFGFW
TMGFRDYNVGSYKIVMLDFDHFENRCQIDVNIEMRKOROPPYFVYGRKSACV
NSIYVSOALIRCKLIAPFLYEEFRDVTPEPSRCERSOQIVLDEPVIATVYTT
SIHMVVKIWSMDAHGTRITYSIRLFSIDMNSIIMNHQIMFWIRPLAVSRGNLFF
HDEKRRL
6308..6472
/rpt_family="BRODYAGAI|BRODYAGAI BRODYAGAI , a
non-autonomous DNA transposon - a consensus."
complement(6311..6472)
/rpt_family="BRODYAGAI|BRODYAGAI BRODYAGAI , a
non-autonomous DNA transposon - a consensus."
complement(<6886..>7206)
/gene="F10F5.5"
complement(6886..7206)
/gene="F10F5.5"
/note="predicted by genemark.hmm"
complement(6886..7206)
/gene="F10F5.5"
/codon_start=1
/product="hypothetical protein; 7206-6886"
/protein_id="ANG51771.1"
/db_xref="GI:12323595"
/translation="MAKILTKFYVFLVLMVSVLLAIEGRPYKDSRSRLTQNRDS
SMFNGSVIMSSFRKPKYDLSWLATVQKSPSGVGHHRKAGYKMGGRANDSGSPS
GVGH"
complement(join(<8349..8423,8505..8799,9001..9309,

9389..9450,9538..9658,9921..10027,10150..10323,
10454..>10573))
/gene="F10F5.6"
/complement(8349..10573)
/note="similar to protein contained within GB:AE001273
from [Chlamydia trachomatis]
complement(join(8349..8423,8505..8799,9001..9309,
9389..9450,9538..9658,9921..10027,10150..10323,
10454..10573))
/gene="F10F5.6"
/codon_start=1
/product="Na+/H+ antiporter, putative; 10573-8349"
/protein_id="AAG51773.1"
/db_xref="GI:12323597"
/translation="MALLFGIGYVLIIFEEISLFSKSGIALLMAYSLWVRSIETSE
IVTLEIHAATSEVSOIVFYMLGAMTIVEIIDAHOGEKLVTDCTISRKPKILLWYGF
TEFLSYVLDITSTIVWYSILRLIPSEYKRLGAVYVIAANAGANTPIGDYTTM
LWIGHISTESTIKNLELPKSLISLVPLALMSLSEVGMGLNTPPTPLADRSAPR
GKLVGEGFGLVPLPKSLTGLPVGILLGIVLWILTDVLIHGDLEKQLKLP
ALSLRDSOGALFFGILLMSLDAAGILKIVANILDAHINVELIASIIGVSAIID
NPLVAAITMGVLDISTEPDSEFWOLISFCAGTGSMLITGSAGVIFMSMEKVFEW
YFRVSGFAFGFTAGIMTYLVAHNPL"
join(11275..11532,11822..12131,12213..12387,12469..12670,
12773..12987,13075..13425)
/gene="F10F5.1"
11275..13425
/gene="F10F5.1"
join(11341..11532,11822..12131,12213..12387,12469..12670,
12773..12987,13075..13425)
/gene="F10F5.1"
/codon_start=1
/product="unknown protein; 11341-13243"
/protein_id="AAG51775.1"
/db_xref="GI:12323599"
/translation="MSFEETPEPLNEKSLVDYIKSPALSSKIGADKSDDLVIKVECG
GNLNFVPIVSSGSIVYKQALPYIRCGESWPMPTKRAIEATTLTKRHGSLSDHP
EYVFDRTMALIGKRYLEPPHILIRKGLIAGIEYFLADHSDYMAKLLFTSLYHD
TTERHRAVTEFCGVELCRLEQVYSDPYVSTFNRTSYLDLDAKAVEDSLKL
EIALKSMFCERAOALIHGDLHTGVMYQDSTQVIDEFSFGPMGDIGAYGNLI
LAFLADGHAIOENDRKREYKQWILRTIEQWNLNPKRPIALMDONKDDPGAYADIY
NNTVEIKVEQENYRNLLHDSLFGAGAKMIRIYGVAVHVEDFESIEEDKRAICERSA
LEFAKMLKEKRRKSTIGEYVSAIOOOS"
complement(join(13447..13659,14434..14535,14726..>14995))
/gene="F10F5.7"
complement(13447..14995)
/note="predicted by genemark hmms"
complement(join(13447..13659,14434..14535,14726..14995))
/gene="F10F5.7"
/codon_start=1
/product="hypothetical protein; 14995-13447"
/protein_id="AAG51776.1"
/db_xref="GI:12323600"
/translation="MRGTDEEDGVNMFSDVLMASDVLSGVEEDSFGRRFTGNRA
PQMLCEGGMNDITDLFSPQSPSPASPLSRNACVYNDSPFSNKKTGADGDHRY
CDKPKGRKRDQKSSLGNAKDDDEGVEVPMKVELRSNGCLVPLAWTVAVANT
NGADLMSALISPLTSHNSEPVEIVEKSHDK"
18149..18708
/rpt_family="novel repeat related to GIB/AB000094, inorganic
phosphate transporter pseudogene. Matches fragments of the
ARNOLDY non-autonomous DNA-transposon family."
complement(18667..18732)
/rpt_family="ATREP21ATREP2 ATREP2 dispersed repeat - a
consensus."
join(<20579..21088,21440..21784,21864..>22493)
/gene="F10F5.2"
20579..22493
/gene="F10F5.2"
join(20579..21088,21440..21784,21864..22493)
/gene="F10F5.2"
/codon_start=1
/product="unknown protein; 20579-22493"

/protein_id="AAG51777.1"
/db_xref="GI:12323601"
/translation="MEKKOGFPSSLRDEYVGLISPSRSRPSRSASPSRSRTTPHKKAL
LMGRKKLIASSGCGGCGVIGERGYYLSQDEPFTIGSSSLRPMEGPPDDNGEYSG
VDSKRLGSLHWYVGWMSRAPSVSTIPATRKSDRLLVGWAQPLAPINWSSSHL
LHLTRDSPTESSAOYILQOYTAAGGKHLHMAIKNAVAGKLMKMTSELETPTGV
RNKSTKSETGTFVLMQNPDMWVVELSVGSGSKVACNGKLVNHRHFWLSHTAKP
VRPLRALOGDIPRTTATMFAKSCVGRKNGEGCFILCTOPELTPARSEGAEL
VRHILFYGESRGTGLAQIEDSOLTRIOSNDGAVVWETTTNSIDYKKOYEGITAH
SGRSVYTLFRRGEVAMSHTRTKMEERMTIEEVARVNPBLSIDCFPRADLSSGSLTEA
CEYSGQEKKGKSSIALASTTRAKRAKVALERGSFNDPWHIDV"
complement(join(<23662..24007,24285..24472,24568..>24780))
/gene="F10F5.8"
complement(23662..24780)
/gene="F10F5.8"
/note="identical to GI:3790593 from [Arabidopsis thaliana]
(FEBS Lett. 436 (2), 283-287 (1998))"
complement(join(23662..24007,24285..24472,24568..24780))
/gene="F10F5.8"
/codon_start=1
/product="RING-H2 finger protein RHYA; 24780-23662"
/protein_id="AAG51778.1"
/db_xref="GI:12323602"
/translation="WTSASLEFSTRRSRPGSRSDPALESSTSYRHHSHHRRGCVH
HNQRHSDGCDPLRRPPTPLRPFHPHRSRPIIDVOGTSOYLTNDSTDTQSSFY
NLNGSERLPAGVILLARDLFERLGRVSLSSRSNRSLIDORESSFSIDGPIFOL
AGLOVTECNKRKPGQIDALINCLHROFSSAEVKSSEWRDSCITLESFTKGMILSLP
CTHSFHSCLNPMWLRACDCCCRALAKE"

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-09-697-089-2 x AC079674 ..

Align seg 1/1 to: AC079674 from: 1 to: 48940

139 LeuTPArGLysaSPGlnHISHis 147

|||||

43196 CTTTGAGCAGACATCAATCATCAT 43222

seq_name: gb_pr:AC083883

seq_documentation_block:

LOCUS AC083883 6019 bp DNA PRI 03-JUL-2001

DEFINITION Homo sapiens clone Rpl1-792N18, complete sequence.

ACCESSION AC083883
VERSION AC083883.6 GI:14589790

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 60199)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished
2 (bases 1 to 60199)
AUTHORS Waterston,R.H.

REFERENCE Direct Submission
TITLE Submitted (04-OCT-2000) Genome Sequencing Center, Washington

JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 3 (bases 1 to 60199)
AUTHORS Waterston,R.H.

TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT On Jul 3, 2001 this sequence version replaced gi:14389375.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0792N18
-----
FEATURES
source
1. 60199
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P11-792N18"
BASE COUNT 16718 a 11679 c 12167 g 19635 t
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment block:
US-09-697-089-2 x AC083883/rev ..
Align seg 1/1 to reverse of: AC083883 from: 1 to: 60199
438 LYSTYLSPHEPHEHISLYSSEPH 446
|||||
37531 AATATTAATTTTCATAGTCCTT 37505
seq_name: gb_pl:AB007645
seq documentation block:
LOCUS AB007645 72698 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K8K14.
ACCESSION AB007645 BA000015
VERSION AB007645.1 GI:2564045
KEYWORDS
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui TAC
clone:K8K14.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis
1 (sties)
Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 5. III.
Sequence features of the regions of 1,191,918 bp covered by
seventeen physically assigned pl clones
DNA Res. 4 (6), 401-414 (1997)
98162728
2 (bases 1 to 72698)
Nakamura,Y.
Direct Submision
Submitted (03-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamukazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=K8K14
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(Informatics Group, Oak Ridge National Laboratory,
http://combio.ornl.gov/Graal-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli.ni.zool.iastate.edu/cgi-bin/sp.cgi).

```

```

Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K8G17 and the 3' clone is K919.
Location/Qualifiers
1. 72698
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="K8K14"
/clone.lib="Mitsui TAC"
complement(join(283..653,770..1619))
/note="contains similarity to unknown protein
gene_id:K8K14.1
sp|Q10058"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09014.1"
/db_xref="GI:9758428"
/translation="MAVSSLPLPPVLTPTTTCSSNSKSAARKQIGETSKRVFV
CGGVIQVCTAYFAFLAKKGIATLVLEKSAVACAAGAGFLAFMCDGSPVAHLARAS
FKLHRSLAEELNGVESYGYRALTTLSVTVEESKPGSGGLPDWVNGPAKSTIGITV
OTTAQVHPOLFTKRLSTAREKYGVEVVGKLEEVESGRVNSVYLEGRIIDVYV
VLAMGWSKSFELLSIFRVCYKASITVEKREPATIPALFTLRPAHGCALDP
EYPRPTGEYITCGNSSQOEVEDDADQVTSNPESTIOYLKRAKTVSSYINENALYKA
EQACFLPSTEDGVPVIGELPIGKGVYGHSCWGLTNGPATGALAEILVDGVAITSV
DLRFSPSRFSKR"
5973..6890
/note="gene_id:K8K14.2
pir|IS71284"
/codon_start=1
/evidence=not_experimental
/product="myb-related protein, 33.3k"
/protein_id="BAB09015.1"
/db_xref="GI:9758429"
/translation="MADRIRKQPMSPEDRDLRIYVYKGPMMVTYISKIPGRSGSC
RIKRCNQLSPQVEHRFSAEDEETIRARAQGNKKATARIKLTNGTDAAVNNHNSI
LKRCGYDRHGDGSEDIRPVKRSVASPPVYGLYSPGSPGSDVSDSTIPIL
PSYELERKVPVRGAVVLPPIETSSSDPPSLSLSLGAVSESSNSHSTNN
TSSRHNNHTVSMFPGSGFALIEEMKSPFGNGCEFMVAQVEMIKAEVRSYTEM
ORNNGGFVGFDNCGMIPMSOIGVRIE"
complement(join(7177..7782,7871..8755))
/note="gene_id:K8K14.3"
/codon_start=1
/evidence=not_experimental
/product="cytochrome P450"
/protein_id="BAB09016.1"
/db_xref="GI:9758430"
/translation="NGEMIDPLALMAGFIYAVYFRKQKKNPNNVYGPVYICHT
HLKEPVHRSKLDLNRNLGIDVILFKLSRRVAVVTSASAEESISQNDVAFANPL
ATLTVMYGNNTLVSTAPYGEHMRRLRRCADVILSTARLRDESDIRREVRAMARKI
NYELVTSQSVRLKLOPLVGLTYNLIMSMVAGKREDETEKVRLLREVDVFGVY
YVGDPLPTLKLDLDGYRRKRAKLSKGLPQKLVDERKREKRALEKLEMTITRLS
LDSEPECYTDIDIGLNOVMIAGDPTAVLLENAMALLHPEVLRKLEINLQVS
KGRVPEESDTGKPLNNVISETLRLFAAFLVPHASSTCEVAGFDIPRTMLFT
LLOCFEMERDNDVAVDMSEKGLTPKRSVLVAKKSLPILDKLV"
join(9132..9192,9618..9763,9851..10349,10429..10523,
10630..10676,10761..10911,11066..11129,11503..11661,
11749..11870,11983..12057,12141..12194,12290..12363,
12464..12538,12619..12838)
/note="gene_id:K8K14.4
ref|NP_005638.1
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09017.1"
/db_xref="GI:9758431"
/translation="MSLTSVSLNPLVFRVLEDSGTTAAFTLGYEAGINKNSIDGM
VPGALIKFVQKGLQYMEAMLSNSEVDDIDDFSFQPLDLSKDVRELQMLREK

```

[illegible]

```

/evidence-not_experimental
/product="natural resistance-associated macrophage
protein"

```

```

/protein_id="BAB09018.1"
/db_xref="GI:9758432"
/translation="MSEIDRRERPLLASERAEYEETKYLIVGIDEEEDADYDDPGNS
PRESKKTLMTLTGPEFLMSIAFLDPGNLESDQALAGSLINLMLMAWAIGLLIOL
LSARIGVATGRHLELCREEEPTWABRWMLMAEALAGADICQVISAIAIKLISNG
LIVLAGVITATLDCFIPLFLENGIRKLEAVPILITATMAAPAKMGEORPGSELT
LVGALVPLTSRRTKOAVGIVGCIIMPNVNLHSAIUSREVPDRKPRKAEALKYYS
IESGTALAVSTINPVYTVFAKSEFTGETADTGLNAGQYLODKTGGCGFFPLIYIM
AIGVLAACOSSTITGTAGQPIWEGFNLKKMKKRAKSIIPMVAIVFDSSSS
DSIMDELEENINLOSQYIPFAVPIPLCTVSNEDIMSGFRIQPIQVTSIVLVAIVLA
INGYLMVDFEFGAATNLLIPLVIFIATAYVFPVLIVLISRLFTVTPQVLAASSHQDQ
JohIn(16968..17243,17316..17432,17512..19215)

```

```
strong similarity to unknown protein"/codon_start=1
```

/evidence-not_experiment1
 /protein_id="BAB09019.1"
 /db_xref="GI:9758433"
 /translation="MEVSYRLVLDLNISSYISLSMMDLSNPAPHKYYTRKEDIKILL
 KVELENLIDSDAPSELNNQFEELAQVVDLRQFQSQWPLSTRIFYVLRTESLAKR
 LRESSLVEVFLKHCHEQLPADLILSPSECEILEYKLTARDEISYITDQALDKQCV
 GPTSDYKIAESTLISNGOELIVGEVYLTNMMKDEALTDQDEIAEYLDGLISLTQM
 HEYLDIDQAOALRCFVRVSPDFCSLSLELTQDPIVYASQGTFFERFLQKIDMGWLM
 CPTKTDASHHTLLPNFLIVRAFLASWCSTNNVYPPDLEILHSESPPELVESYRASS
 SENGSESLDDELQVYSRSASAPGIYSEVYCCYKTRNNNAADRLSTRSNTPKFPFE
 ERMHNPDIPIATVREITGSSSIETEVKRLIDDLCTSSLYDQKATARIILANLNSD
 NRIYVAREAPISVLISLYSTDERIQDAVATCLNLSINDNNKSLIAESGAIYPLIHV
 LKTYGAEKAKNSAATLELSLISYIEYKTEIECAAIPELYLDLSSGSLCKADQATATL
 ENLSLGHENKTRKVIAGAVRYLVLEIMEDQAPFQWMAVYLANLTAVREGKALIEEGEG
 IPIVLEYVELISARGKEVATALLDQCTHSPKPCNNYIRGEVITPEYVALTKRSGAROK
 EKVYLFPLFLCYVNS"
 complement(20385...21293)

CDS
 EVKYLFPLLLCIVANS"
 complement(20385..21293)
 /note="gene_id:K8K14.7
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09020.1"
 /db_xref="GI:9758434"
 /translation="MAMTETDEDKTLILISRYVEDDDVEEEEEEEALSLCDLPNEK
 GELRLIVEEDDEDFSGFEFGIGSSFRAGSDSCPAPEMSTADELFKGRILPLRHVS
 SLDALNPETRLITRSESEVERRTGIIISDKIKNNFTDYQSPQPIRRSSMTKAT
 VNSINPPSSSIIMDLRLGLVTRPEIETLTAGNAKLSVNSSCSSITNSNNKKG
 SGESNRNRRRRSPMSDDCKCSVSTETKAPKIVNVSSETEERKMMKTKAKEENS
 AMAKRRTLEMLSQVGEFVYDGRRTLY"
 complement(24056..26329)
 CDS

CDS
complement(24056..26329)
/note="gb|AAC18851.1
gene_id:K8K14.8"
/codon_start=1
/evidence=not_experimental
/product="Cucumisin-like serine protease"
/protein_id="BAB09021.1"
/db_xref="GI:9758435"
/translation="MSSSFLSTAFELLCIPGHVSSSSDGGTYIVHAKSMPSS
EDLHNKWDSSLRISDSAEILYYENKIHGFSTRLLQEPKDSLMTQPGYISVLPENR
YELHRTLPFGEDENHADIPEAGSSIDVAVGLDTPWPEKSTSDSEFGPISS
WKGGCEAQTINCSNKRLLIGARFANGYSTIGPIDENKSSPPDDHOGHTTSS

TAACSVVEGASLLGYASTGTRGMAPRARVAVVYVCWJLGGCFSSDILIAADIKATADANN
VLSLSLGGGMSDTRYRDCVATIGAFAPAMERGLVILSCSAGNACBPSSSLSNADAPMTTNGA
GTLDRDPALAILANGNKNFTVGLSEKAEALPDKLPIYAGNLSMANWMLCMITGLI
PEKAKGKIVMCDRGINARVOKGDVYKKAAGVGMLIANTANGELVADHILPATTVG
EKADDIRHYVTPTDPNPATASISLIGTVAGVGPSPVVAFTSRGNSITPAILPDLIA
PGVMIILAAWTGACAPTCPLADSDRVEENITISGTSMSCPHYSGLAALLKSVHPEMSRA
IRSLALMTTATKTYKIDGKPLDLIDATGKSTPFDGAGVSPPTATNPCLITDITDEL
GFLCALNYSPOIRSRSRMYTCDPSKSVADILNPAVNDVGAVTGYRTFVSV
GGAATYSVKVTSFETTGKIVSEPAVLNKEANEKKSVTYFTVDSSKPSGSSNSFGSIE
WSDKHVGSPPVATSWT"

complement({oin(29531..29693,29772..29953,30043..30078,
30480..30575,30665..30742,30830..30982,31054..31329))
/note="gb|AAC18972.1
gene_id:K0K14.9

```
alignment_scores:
  Quality: 9.00
  Ratio: 1.000
  Gaps: 0
Percent Similarity: 100.000
Percent Identity: 100.0000
```

```
alignment_block:
US-09-6997-089-2 x AB007645 .
```

Align seg 1/1 to: AB007645 from: 1 to: 72698

239 LeuArgGlnArgValLeuPheLeu 247

39915 CTCGGGCAAGAGTCTTTTCTCTT 39941

seq_name: gb_pr:AC002122

```
seq_documentation_block:
  locs: AC003122
```

DEFINITION	Human unknown clone GS1-293J4 from 5p15.2, complete sequence.
ACCESSION	AC003133

VERSION AC002122.1 GI:2121322
KEYWORDS HTG

SOURCE human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo

REFERENCE	1 (bases 1 to 79931)
ALTMAN, J. A. and Erickson, B. J.	1990, <i>Proc. Natl. Acad. Sci. USA</i> , 87 , 1000-1004.

TITLE The sequence of *H. sapiens* unknown clone GS1-293J4

REFERENCE	2 (bases 1 to 79931)
ALTHORS	Waterston P

TITLE	Direct Submission
JOURNAL	Submitted 721-MAY-1997

REFERENCE	3 (bases 1 to 79931)
AUTHORS	Waterston, R.

TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 79931)

AUTHORS Waterston, R.
TITLE Direct Submission

JOURNAL
Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center

Center code: WUGSC
Web site: <http://genome.wustl.edu/qsc>

----- Summary Statistics -----
Contact: sapriens@watson.wustl.edu

Center project name: H_GS293J04

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap

between neighboring data submissions

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics/Genome Sequencing Center,
Washington University School of Medicine.

Mapping information for this clone was also provided by Dr. Michael
Lovett, University of Texas Southwestern Medical Center, Dallas TX.

SOURCE INFORMATION:

This clone is from the first BAC library from Genome Systems, Inc.
(<http://www.genomesystems.com>).

Cell line: lymphoblastoid

Haplotypes: two

VECTOR: pBelotBAC

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is GSI-259010, 200 bp overlap; the
clone sequenced to the right is GSI-11304, 200 bp overlap. Actual
start of this clone is at base position 195 of GSI-293J4; actual
end is at 46016 of GSI-11304.

Location/Qualifiers

FEATURES

source

1. 79931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="GSI-293J4"
/clone_lib="GSBAC1"
/map="5p15.2"
264. 391
/rpt_family="ALU"
complement(524. 1483)
repeat_region
/rpt_family="THR"
complement(1853. 3416)
repeat_region
/rpt_family="THR"
2591. 2608
/rpt_family="LI"
complement(3417. 3530)
repeat_region
/rpt_family="LI"
3534. 3811
/rpt_family="ALU"
complement(3812. 4270)
repeat_region
/rpt_family="THR"
complement(4271. 4341)
repeat_region
/rpt_family="LI"
4368. 4501
/rpt_family="ALU"
5721. 6241
repeat_region
/rpt_family="LI"
8188. 8478
repeat_region
/rpt_family="ALU"
complement(9886. 9939)
repeat_region
/rpt_family="LI"
complement(11042. 12129)
repeat_region
/rpt_family="LI"
12114. 12619
repeat_region
/rpt_family="LI"
complement(13773. 14228)
repeat_region
/rpt_family="LI"
14209. 15600
repeat_region
/rpt_family="LI"
16112. 16402
repeat_region
/rpt_family="ALU"
16496. 16727
repeat_region
/rpt_family="ALU"
complement(18332. 19676)
repeat_region

repeat_region
/rpt_family="LI"
19661. 20405
repeat_region
/rpt_family="LI"
complement(21840. 21896)
repeat_region
/rpt_family="LI"
complement(22471. 22494)
repeat_region
/rpt_family="LI"
complement(24731. 26658)
misc_feature
/note="90% similarity to Human DNA sequence from cosmid
L30G1, Huntington's Disease Region Z68870 (NTD:91164910)"
repeat_region
/rpt_family="LI"
complement(2707. 27753)
repeat_region
/rpt_family="LI"
complement(28006. 34144)
repeat_region
/rpt_family="LI"
28988. 29405
repeat_region
/rpt_family="LI"
complement(36343. 36363)
repeat_region
/rpt_family="LI"
complement(36563. 36612)
repeat_region
/rpt_family="LI"
complement(37450. 37469)
repeat_region
/rpt_family="LI"
complement(37979. 38007)
repeat_region
/rpt_family="LI"
complement(38105. 38762)
misc_feature
/note="90% similar to Human TCB gene encoding cytosolic
thyroid hormone-binding protein M26252 (NTD:9338826)"
repeat_region
complement(38448. 38474)
repeat_region
/rpt_family="LI"
38626. 38662
repeat_region
/rpt_family="LI"
40036. 40329
repeat_region
/rpt_family="ALU"
40867. 41232
repeat_region
/rpt_family="ALU"
complement(41431. 41450)
repeat_region
/rpt_family="LI"
41898. 41936
repeat_region
/rpt_family="LI"
complement(43951. 44007)
repeat_region
/rpt_family="LI"
complement(44620. 44916)
repeat_region
/rpt_family="ALU"
complement(45481. 45506)
repeat_region
/rpt_family="LI"
46103. 46144
repeat_region
/rpt_family="LI"
47513. 47545
repeat_region
/rpt_family="LI"
48923. 49187
repeat_region
/rpt_family="ALU"
complement(50934. 51743)
repeat_region
/rpt_family="LI"
51093. 51511
repeat_region
/rpt_family="LI"
51732. 52559
repeat_region
/rpt_family="LI"
54961. 55100
repeat_region
/rpt_family="LI"
55102. 55378
repeat_region
/rpt_family="ALU"
55380. 55810
repeat_region
/rpt_family="LI"
56159. 56251
repeat_region
/rpt_family="LI"
56424. 56692
repeat_region
/rpt_family="ALU"
58720. 58744
repeat_region
/rpt_family="LI"
complement(59915. 60055)
repeat_region
/rpt_family="LI"
60117. 60322
repeat_region
/rpt_family="ALU"

```

repeat_region      complement(61568..61829)
                    /rpt_family="ALU"
repeat_region      62285..62575
                    /rpt_family="ALU"
repeat_region      62992..63019
                    /rpt_family="L1"
repeat_region      complement(63297..63316)
                    /rpt_family="L1"
repeat_region      64371..64391
                    /rpt_family="L1"
repeat_region      complement(64828..65175)
                    /rpt_family="L1"
repeat_region      complement(65717..67547)

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:
US-09-697-089-2 x AC002122/rev ..

Align seg 1/1 to reverse of: AC002122 from: 1 to: 79931

```

809 AaPTyTtIleVallySerleuserSer 817
|||||
37359 GACTACATTGTCAAAAGCTGTCAACT 37333

```

seq_name: gb_pl:F14D7

```

seq_documentation_block:
LOCUS      F14D7      81513 bp      DNA      28-JUN-2000
DEFINITION Sequence of BAC F14D7 from Arabidopsis thaliana chromosome 1,
complete sequence.
ACCESSION  AC021198
VERSION     AC021198.2  GI:6957696
KEYWORDS    HTG.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 81513)
Liu,S.X., Chan,A., Sakano,H., Yu,G., Lee,J.M., Lenz,C., Pham,P.,
Toriumi,M., Chin,C., Chlou,J., Choi,E., Chung,M., Gonzalez,A.,
Hwang,B., Liu,A., Vaysberg,M., Altafi,H., Brooks,S., Buehler,E.,
Chao,Q., Conn,L., Conway,A.B., Hansen,N.F., Johnson-Hopson,C.,
Khan,S., Kim,C., Lam,B., Miranda,M., Nguyen,M., Palm,C.J.,
Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R., Federspiel,N.A.,
Theologis,A. and Yu,G.
The sequence of BAC F14D7 from Arabidopsis thaliana chromosome 1
unpublished
2 (bases 1 to 81513)
Theologis,A.
Direct Submission
Submitted (15-JUN-2000) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 81513)
Theologis,A.
Direct Submission
Submitted (09-FEB-2000) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 81513)
Theologis,A.
Direct Submission
Submitted (28-JUN-2000) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA

```

COMMENT
On Feb 11, 2000 this sequence version replaced gi:6693723.
The sequence is of BAC F14D7 from Arabidopsis thaliana chromosome
1. The sequence does not represent the sequence of the entire
insert of this clone. It is shorter by 6954 bp because we submit
only the unique sequence of the clone. However, in order to
facilitate the joining of overlapping clones in the future for

FEATURES Source

creation of larger contigs, we provide small overlaps (200 bp)
between overlapping subitted clones. The 5' end of this sequence
overlaps by 200 bp to the 3' end of the sequence of the clone
F1504.

```

FEATURES
Source
Location/Qualifiers
1..81513
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F14D7"
1553..4999
/gene="F14D7.1"
join(1553..4628,4713..4999)
/gene="F14D7.1"

```

/note="Contains similarity to receptor protein kinase-like
protein from Arabidopsis thaliana gb|A161513. It
contains a eukaryotic protein kinase domain PF100069. EST
gb|A1997574 comes from this gene."

/evidence=not_experimental
/protein_id="AAF79881.1"

/db_xref="GI:8778966"

/translation="MGAFAKNDYDFRFLFISITLSCSISASATIAEAMLLKKSTF
TNSKLSWHDANTNTFSCTSWYGVSCNSRGSIEBLNTNGIEGTFODPFISLS
NLAYVDLSMLLSGTLIPPOFNGLSKLIYFSLTNGHLEGTSESGNKNTLYLHON
YLSVISELGNMESMTDLASONKLTGSPSSSGNKNTLVLYENLYTGVLPPEL
GNMESMTDLASONKLTGSPSSSGNKNTLVLYENLYTGVLPPEL
LSONKLTGSPSSSGNKNTLVLYENLYTGVLPPEL
PSSLGNKNTLVLYENLYTGVLPPELGNMESMTDLASONKLTGSPSSSGNKNTLVLYENLYTGVLPPEL
TYLYLVNLYTGVLPPELGNMESMTDLASONKLTGSPSSSGNKNTLVLYENLYTGVLPPEL
SGALPVGANSNHLTTLIDPNTNFTGFPPEYVCKGKRLQNIISDYNHLESPKSLND
CKSLIRARFLGNKFTGDIFFAAGIYPLNFTDFSHKFGCEISNMKSPKGLIWS
NNNTGALPPEINMQLVDELSTNNLPGLPAPALGNLTNLSRLNGNLSGRVPA
GSLFTNLSESLSSNPFSEIROPSPFLKIDMNTSRKPKGSPRLSKLQIOTD
DLSNOLDGEIPSOSSLQSLDKLDSLHNNLSGILPTTFEGMALTVDSNNLCEP
LPDPTFRKATADALEENIGCSNIPKRLPCKELPKKNGLVVWVILPLGLVLY
ILSICANTFYICIRKRLONGRNTDPTGENMSIFSYDGKFEYODIESTNEPDTYL
IGGYSKYVRANLODTIIVAKRLHDTIDEISIPYVQKLEFVKALEFIRHNVK
LFGCSHRHRTFLIYEMEGSLKLANDEAKRLMTKRLNVYKGVANALSHMD
RIRPIVHRDSSGNILLNDYPTAKISPGKAKILKTDSSMWSVAGTYGVAPAEAT
MKYTEKCDVYFSGVILLIILTIKRPGLVSLSSPPEALSLNISIDERVLEPRGQR
EKLLKWEAMLLCLQANPESPTMLISTIFS"

5917..7552

/gene="F14D7.2"

join(5917..5992,6589..6734,6821..7552).

/gene="F14D7.2"

/note="Identical to annexin (AnnA1) mRNA from Arabidopsis
thaliana gb|AF083913. It contains an annexin domain
PF100191. ESTs gb|H76460, gb|I218518, gb|I226190,
gb|IN96455, gb|I247714, gb|I41940, gb|I743657, gb|IN95995,
gb|IR30014, gb|I22046, gb|H73398, gb|H77008, gb|R29768,
gb|H36260, gb|I217514, gb|H43175, gb|I76739, gb|AA112753,
gb|H76134, gb|I742209, gb|H36536, gb|AI998553, gb|I23265,
gb|AA597533, gb|AI100145 and gb|AI10054 come from this
gene."

/codon_start=1

/evidence=not_experimental

/protein_id="AAF79882.1"

/db_xref="GI:8778967"

/translation="MATLKVSQSVAPSDAEDQRTAFEGWGTNEDLIISILAHRSAE
QKRVIRQAYHETGEDLLKTLKELSDPFAIILMTLEGEDRALANATKRWSS
NOVLMVACRTSTQTLHARQAYHARYKSLSEEDVAHHTGGDFKLLVSLVTRYSG
DEVNMTLAKQKAEKLVHEKIKDKHNDVDVIRIISTRKAOINFTFNQDDHGEILK
SLREGDDDDKFLALRSTIOCTPPELYPDVILRSAINKKTGDEGALTRVYTRAEID
LKVIQEYQRKNSIPLKATFKDRGDEKMLVALGEDDA"

complement(7937..10730)

/gene="F14D7.3"

complement(join(7937..8041,8115..8364,8443..8606,
8757..8919,9397..10730))

/gene="F14D7.3"

/note="Contains similarity to a putative protein CAB78009
gb|I7267527 from Arabidopsis thaliana BAC T32A17

gene

CDS

```

gb|A1161513. It contains Pumilio-family RNA binding
domains PF100806."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF99883.1"
/db_xref="GI:8778968"
/translation="MKKNNKRRKGRKSVNSISSAGCKKOCRLCYDINGEKTKENT
TSLRTVYATATFSLSLQSITCYITKRLSPYCFGLTSLSPFCVNLGYSEM
FGRRMGFGGNDENFRGASDRIPGSDGFLPSLDNPELKNKRSVPLDGLCKIAK
MGSCMSITATKREDFRDPSEFGARTLHGSSGFQDNLGASQIHDFRNPSSVQLQ
NNHGVSPSPGMRLLGRQDSFNLNGFEEMALKNHDFLDQIHEPIKRPFLGN
DALISSMEEGNNRVSQTLAMEASRGFPPEEDSSLSPEHVKVSGKLASVLEIVL
IGGSGYKMSPKSNDLVAMEASRGFPPEEDSSLSPEHVKVSGKLASVLEIVL
IDHVELSKDPGNYIYOKLFVSDDEORTLIVSYTSMRELITICNTGSTRVVO
KMETVTKQOALVAKSGKPGFLAVKLDNGHVLVOSCLOTGPRNDNEVLEATKY
CAEIAIHRHGCCVLOCCISNSVGLQREBLVAELSRSLHSQDPFRNYVOYLIDQV
SAVKLVQFRMHVIAELATQFSSHVIEKCLRKIPESRAELVRLCLVPMVEYLLDPPY
ANVYIQTLAVTKGPVRAKLVAKVRYRKLHSSPYCKIKITILKK"
12116..14526
/notes="Contains similarity to Tnl-1 transposon gb|X13291"
complement(19422..23686)
/gene="F14D7.4"
complement(join(19422..19496,19746..19835,20053..20194,
23653..23686))
/gene="F14D7.4"
/notes="Contains similarity to 8-amino-7-oxononate
synthase-like protein gi17406466 from Arabidopsis thaliana
BAC T32M21 gb|A1162875. EST gb|A1994915 comes from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF99884.1"
/db_xref="GI:8778969"
/translation="MAKKATVALICFTLGEELFSGVFAECRGRKKLLSVNDYL
GISSHTISNINANLCEEGILGCCSTSTFSLKILKHAFTLEKVDADDERKKNED
EYGEDDDYKR"
complement(34011..36183)
/gene="F14D7.5"
complement(join(34011..34115,34201..34450,34532..34788,
34873..35035,35279..36183))
/notes="Contains similarity to a putative RNA-binding
protein AAF18550 gi16587864 from Arabidopsis thaliana BAC
T1111 gb|AC012680. It contains Pumilio-family RNA
binding domains PF100806."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF99885.1"
/db_xref="GI:8778970"
/translation="MEIFNFGASDHRRLPDGSGGFLQSLDTPNPLKNQYNNNSVEA
LECKRLNKGISCDMSITKPEPRVDPGCGAKTLHESFGFQNLGASQIHDFG
RNPSSRYVONNNFHCYSPGCLIGDQDSFNPNGFEEMAFKHDKDLIHINEPIKRS
PFLRGDADKSGSLMEGIVSQTLAMEASGASYPDEPKINSGLPLDVSWEITGVS
NIMARQDQICRALQKLEGEVTSQILMEIDHVELSMPLGNVLYOKLIVASDE
EORTMIVSVLSTCELEFLPNDKRLTKLHIIISPLFLTVYSSDDCCICWQFVVEA
ATEYCAQLATHOVCYVLOCCISLIMYGLQREBLVAELSRSLHSQDPFRNYVOYL
DOOVSVNLLPFRHICELATQFSSHVIEKCLRKIPESRAELVRLCLVPMVEYLL
DDPYANVYIQTLAVTKGPVRAKLVAKVRYRKLHSSPYCKIKITILKK"
44612..48698
/gene="F14D7.6"
join(44612..44780,45014..45403,45501..45644,45676..46134,
46297..46476,46633..47029,47113..47505,47737..48082,
48145..48411,48465..48541,48632..48698)
/notes="Strong similarity to a mutator-like transposase
gi14063759 from Arabidopsis thaliana BAC T14A4
gb|AC005561."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF99886.1"
/db_xref="GI:8778971"
/translation="MSDIYGVARHLYTLVLLAVYSLCIKEKPTRLKTLVSDKINA

```

```

KCAPHENSRLKGVRLREODGFAYVMAVAKELALGAEDEDIETTYQPMQMMGPDMR
ANFDILDEDEMTLPMATRADIDEVHPRKTIIRGNSRVNSRSLDGLIEETIS
DKYNSAETRAYWGTALTRMLRNISAHLSPTTYXATITLYQTVMMQPRERTTA
WARRKAMTNETLMGGAGTSGVPLTDEPDSIPISGKRVTDDEDERKLEEMRT
OQOEEMWREMCATQMTLGATQIDANTVGGGTATRALNGSDIEFGNSREDQLTLA
RAETQEVCTQVVALSSSDSTFGMPSTLSYSPSSDLDVDYDSSNNSSCIDKKG
IMDERITTRKEIGAGSTNTRVSEVLEAMVKOTPPVTLVORDADPYDDPEKDAL
HRAKXVADNEGDDITYGRIEKSDECATKLAIHAKRFHPIYAKSGPPIVAVCYSD
TCBPWRYARKLSDSMEFVRSSTQHTCSVDARQDPRKASVYICKLMRTYIGYGR
GPRANERMLRQDVALIOPFERCLETNPNSLVAMETKDNDSVERKYLFALDAC
VQGYAVMRVVIYIDGTHLGRGCGCLVASADANQVPIAFIGVYNSNDRAWTFM
TKLEALPDDELVEFASDRHNSTITPAVRIFVEVRAKHGBCADYLEGIEFHTSHF

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
us-09-697-089-2 x F14D7

Align seg 1/1 to: F14D7 from: 1 to: 81513

758 SerleungIyAsnLeuTyAsnLeuThr 766
|||||
1985 TCCTTAGGAATCTCAGACTTGACG 2011

seq.name: gb-pl:ATT16L4

seq_documentation_block:
LOCUS ATT16L4 88011 bp DNA PLN 17-JUN-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T16L4 (ESSA
project).
ACCESSION AL079344
VERSION AL079344.1 GI:5123543
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 88011)
Bevan,M., Rose,M., Hempel,S., Entlan,K.-D., Bancroft,I.,
Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schellier,C.
unpublished
2 (bases 1 to 88011)
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
COMMENT
unpublished
2 (bases 1 to 88011)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (17-JUN-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de,Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
Location/Qualifiers
1..88011
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
47..1157
/gene="T16L4.10"
47..117
/gene="T16L4.10"
/number=1
join(47..117,312..365,446..551,648..716,803..946,
1056..1157)
/gene="T16L4.10"
/notes="similarity to X-Pro dipeptidase, Homo sapiens,
PIR2:A32454"

```

```

/codon_start=1
/product="X-Pro dipeptidase-like protein (fragment)"
/protein_id="CAB45310.1"
/db_xref="GI:5123544"
/translat="AVLDAHNSVISAARKPGVNVDMHKLAEKLTILESLLKGSILITGDV
DDMWORLIGLAVEMPHGLIGHMGIIDTHDGGYRPGVSRPKRGLKSLRTADLLLEGVYI
TYVRCGYFLKALLFEPAMANAATTSKFFNREIIEFRNPGVRIESDLVYIANGCKNMNTN
VRETIWEIEVWAGGWPPTK"
intron
118. .311
/gene="T16L4.10"
/number=1
312. .365
/gene="T16L4.10"
/number=2
366. .445
/gene="T16L4.10"
/number=2
446. .551
/gene="T16L4.10"
/number=3
552. .647
/gene="T16L4.10"
/number=3
648. .716
/gene="T16L4.10"
/number=4
717. .802
/gene="T16L4.10"
/number=4
803. .946
/gene="T16L4.10"
/number=5
947. .1055
/gene="T16L4.10"
/number=5
1056. .1157
/gene="T16L4.10"
/number=6
1157. .1575
/gene="T16L4.20"
1575. .1812
/gene="T16L4.20"
/number=1
1812.1908. .1918.2041. .2196.2293. .2356.
2437. .2579.2670. .2757.2865. .2980.3071. .3221.3383. .3588)
/gene="T16L4.20"
/number=1
1812.1908. .1918.2041. .2196.2293. .2356.
2437. .2579.2670. .2757.2865. .2980.3071. .3221.3383. .3588)
/notes="contains EST gb:H37349, T45910, H37681, Z26403,
AA041059"
/codon_start=1
/product="arginine methyltransferase (pam1)"
/protein_id="CAB45311.1"
/db_xref="GI:5123545"
/translat="MTKNSNHDNEFIISPEPNQTKIRPEDADEDEVAEGSGVAGBET
PDESEFMDGESADRAEYVDITTSADYTRDYSNHSHEGIEHMLKDVYRRTTYQAVIYN
KFLDKYILVDAGAGTILSLFCARGAHAHYAVECSDADAKETLVKANGFSDVITV
LNGKLEELIETLPKVDVYIISEMNGYFLFENMLDSVLYARDKLVYGGVLLPDKASLIH
LTAISEKSEKEDIEFWSNVYCFDMSCKIRKAMBEIYDVTQDNQIATFDSRLIKTMDI
SKMSSDSAFETKAPFKLYAORNDY IHALVAYFDVSPFMCHKILGFSFGPKSRATHMKT
VYILEDVLTICGEFTITGTMSVSPNKKNPRIIDIKLSTSLNGHCKISRQNHKKR"
intron
1813. .1907
/gene="T16L4.20"
/number=1
1908. .1918
/gene="T16L4.20"
/number=2
1919. .2040
/gene="T16L4.20"
/number=2
2041. .2196
/gene="T16L4.20"
/number=3
2197. .2292
/gene="T16L4.20"

```

exon	/number=3 2293. .2356 /gene="T16L4.20" /number=4 2357. .2436 /gene="T16L4.20"
intron	/number=4 2437. .2579 /gene="T16L4.20" /number=5 2580. .2669 /gene="T16L4.20"
intron	/number=5 2670. .2757 /gene="T16L4.20" /number=6 2758. .2864 /gene="T16L4.20"
exon	/number=6 2865. .2980 /gene="T16L4.20" /number=7 2981. .3070 /gene="T16L4.20"
intron	/number=7 3071. .3221 /gene="T16L4.20" /number=8 3222. .3382 /gene="T16L4.20"
intron	/number=8 3383. .3588 /gene="T16L4.20" /number=9 3589. .3892 /gene="T16L4.30"
exon	complement(3892. .4080) /gene="T16L4.30" /number=1 3892. .5493 /gene="T16L4.30"
gene	complement(join(3892. .4080,4202. .4333,4424. .4501, 4614. .4745,4855. .4923,5083. .5193,5284. .5493)) /gene="T16L4.30"
CDS	complement(join(3892. .4080,4202. .4333,4424. .4501, 4614. .4745,4855. .4923,5083. .5193,5284. .5493)) /note="contains EST gb:H76475" /codon_start=1 /product="putative protein" /protein_id="CAB45312.1" /db_xref="GI:5123546" /translation="MVVMNRYLVALTLLALSSALLPVSDAKKSPSTPRKEDPYLIK GVNKLSSRLHQLVKEKQOQISPKTISEYEIIETFAENVCLKKEADMDKIIDIEYK GNYLVIVSQDEBGMNSCKRTENACQKVIQSPDVAEYIYKSPDLVSLVNHCLCD LTDACSRRPPYKDFVGEPPVAKPSKDAEMDITLSMOGMPGAPGMKYISREDIEK AGVSNRYRVMWKGGSSSSKKPKSGKSEL" complement(4081. .4201) /number=1 complement(4202. .4333) /gene="T16L4.30"
exon	/number=2 complement(4334. .4423) /number=2 complement(4424. .4501) /gene="T16L4.30"
intron	/number=3 complement(4502. .4613) /number=3 complement(4614. .4745) /gene="T16L4.30"
exon	/number=4 complement(4746. .4854) /number=4

exon complement(4855..4923)
/gene="T16L4.30"
/number=5
Intron complement(4924..5082)
/number=5
exon complement(5083..5193)
/gene="T16L4.30"
/number=6
Intron complement(5194..5283)
/number=6
exon complement(5284..5493)
/number=6

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x ATT16L4 ..

Align seg 1/1 to: ATT16L4 from: 1 to: 88011

762 Leulysasnleuthrlyslleuilemet 770
|||||
63207 TTAAAAAATCTTACAAACTAATATATG 63233

seq_name: gb_pr:HS1189K21

seq_documentation_block:
LOCUS HS1189K21 100625 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 1189K21 on chromosome Xg26.3-27.3.
Contains two pseudogenes similar to MADH-Ubiquinone Oxidoreductase chain 1 and chain 2 (EC 1.6.5.3), STS and a GSS, complete sequence.
ACCESSION AL030997 GI:3355873
VERSION AL030997.1
KEYWORDS HTG; MADH-Ubiquinone Oxidoreductase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 100625)
Grafham,D.
Direct Submission
Submitted (04-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 30, 1998 this sequence version replaced gi:3288419.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 1189K21. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true right end of clone dJ331L4 (AL022719) is at 100 in this sequence.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChX>
1189K21 is from the library RPEC5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES
Location/Qualifiers

source
1..100625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q26.3-27.3"
/clone="RP5-1189K21"
/clone.lib="RPEC5"
2190..2376
/note="MIR repeat: matches 8..208 of consensus"
2540..2941
/note="MSTB repeat: matches 1..403 of consensus"
2804..2979
/note="MSTD repeat: matches 233..394 of consensus"
complement(1308..3494)
/note="L1PB3 repeat: matches 895..505 of consensus"
complement(3476..3900)
/note="L1PB2 repeat: matches 425..1 of consensus"
complement(3752..7739)
/note="L1 repeat: matches 5390..1401 of consensus"
complement(7705..7751)
/note="MER25 repeat: matches 1559..1513 of consensus"
complement(8478..8683)
/note="MER25 repeat: matches 1510..1305 of consensus"
9976..10271
/note="AluY repeat: matches 2..297 of consensus"
complement(10350..11155)
/note="L1MA6 repeat: matches 1043..240 of consensus"
11567..11598
/note="16 copies 2 mer tg 100% conserved"
11912..11991
/note="MER4C repeat: matches 22..101 of consensus"
12082..12133
/note="MER4A repeat: matches 136..186 of consensus"
12379..12563
/note="MER4C repeat: matches 288..472 of consensus"
complement(12568..12725)
/note="L1MB3 repeat: matches 215..51 of consensus"
complement(12718..13016)
/note="L1PB8 repeat: matches 910..610 of consensus"
complement(13022..13320)
/note="AluX repeat: matches 297..1 of consensus"
complement(13321..13561)
/note="L1 repeat: matches 5319..5066 of consensus"
complement(13554..13727)
/note="L1PB5 repeat: matches 627..454 of consensus"
complement(13730..14827)
/note="L1 repeat: matches 4176..3054 of consensus"
14837..15191
/note="RHE1B repeat: matches 1..364 of consensus"
complement(15434..15946)
/note="L1PA12 repeat: matches 911..390 of consensus"
15947..16304
/note="RHE1C repeat: matches 3..370 of consensus"
16306..16734
/note="RHE1B-INTERNAL repeat: matches 1..427 of consensus"
17597..17761
/note="match: STS L4123"
19153..19285
/note="MIR2 repeat: matches 1..137 of consensus"
complement(19327..19483)
/note="MIR repeat: matches 207..47 of consensus"
19679..19718
/note="MIR repeat: matches 106..145 of consensus"
complement(22100..22156)
/note="MADE1 repeat: matches 80..1 of consensus"
22621..22923
/note="AluSg repeat: matches 1..303 of consensus"
23533..24461
/gene="dJ1189K21.1"
23533..24461
/gene="dJ1189K21.1"
/note="pseudogene similar to MADH-Ubiquinone Oxidoreductase chain 1 (EC 1.6.5.3); match: proteins

```

096189 000505 P92475 P24967 P03887 P48652 P92604 003850
P48900 P38598"
/codon_start=1
/pseudo
/evidence=not_experimental
repeat_region 24621..24981
/gene="THE1B repeat: matches 1..364 of consensus"
gene 25034..25722
/gene="dJ1189K21.2"
25034..25722
/gene="dJ1189K21.2"
/pseudo
/evidence=not_experimental
Oxidoreductase chain 2 (EC 1.6.5.3); match: proteins
095704 Q34799 P03891 P92691 Q03197 Q96061 P38599 Q36346
Q21326 P24970 Q00540 P48653 P41297 P03892 P92476 Q21211
P48905 Q21129 Q21184 Q21305 Q21124 Q21676 P03893 Q21215
Q21303"
/codon_start=1
/pseudo
/evidence=not_experimental
26169..26204
/note="18 copies 2 mer tg 92% conserved"
repeat_region complement(27967..28091)
/note="L1Pa2 repeat: matches 902..777 of consensus"
28394..28451
/note="MIR repeat: matches 119..176 of consensus"
repeat_region complement(28913..28971)
/note="MIR2 repeat: matches 141..83 of consensus"
29036..29238
/note="MIR repeat: matches 53..261 of consensus"
repeat_region complement(29934..30046)
/note="L1Ma7 repeat: matches 1036..924 of consensus"
30053..30335
/note="L1Pa3 repeat: matches 598..883 of consensus"
repeat_region complement(30337..30417)
/note="Alusg repeat: matches 81..2 of consensus"
repeat_region complement(30470..30752)
/note="L1Ma7 repeat: matches 935..647 of consensus"
repeat_region complement(30704..31382)
/note="L1Pa7 repeat: matches 890..210 of consensus"
repeat_region complement(31380..33133)
/note="L1 repeat: matches 4677..2924 of consensus"
repeat_region complement(33308..33617)
/note="L1 repeat: matches 335..12 of consensus"
repeat_region 34444..34743
/note="Alusg repeat: matches 1..301 of consensus"
repeat_region complement(34889..35405)
/note="L1Ma4 repeat: matches 1042..518 of consensus"
repeat_region complement(35430..36327)
/note="L1Pa14 repeat: matches 895..1 of consensus"
repeat_region complement(36176..36411)
/note="L1 repeat: matches 5390..5158 of consensus"
repeat_region complement(37826..38120)
/note="Alusx repeat: matches 302..1 of consensus"
repeat_region complement(39125..39262)
/note="MIR repeat: matches 255..117 of consensus"
39924..40225
/note="Alusx repeat: matches 1..302 of consensus"
repeat_region complement(40555..40791)
/note="MIR repeat: matches 250..1 of consensus"
41635..42113
/note="MLT2D repeat: matches 1..459 of consensus"
repeat_region 42067..42134
/note="MLT2A repeat: matches 307..374 of consensus"
47844..48053
/note="MER2C repeat: matches 1338..1537 of consensus"
repeat_region complement(48470..48581)
/note="L1Pa2 repeat: matches 893..782 of consensus"
48576..48814
/note="L1Pa5 repeat: matches 454..692 of consensus"
repeat_region 49335..49637
/note="AluY repeat: matches 1..301 of consensus"
50526..50559

```

```

/note="17 copies 2 mer tt 82% conserved"
complement(50564..50865)
/note="AluY repeat: matches 300..1 of consensus"
misc_feature complement(51457..51713)
/note="match: STS L24620"
repeat_region 51716..51990
/note="MER31 repeat: matches 60..329 of consensus"
repeat_region 52042..52067
/note="13 copies 2 mer tg 92% conserved"
repeat_region 52734..53533

alignment_scores:
  quality: 9.00      length: 9
  ratio: 1.000      gaps: 0
Percent similarity: 100.000  Percent identity: 100.000

alignment block:
US-09-697-089-2 x HS1189K21 ..

Align seg 1/1 to: HS1189K21 from: 1 to: 100625
848 ValysLeuSerIleLeuAspLeuSer 856
|||||
33005 GTTAAGTTGTCATTTAGATCTTCC 33031

seq_name: gb_pr:AP001576

seq_documentation_block:
LOCUS AP001576 105199 bp DNA PRI 16-AUG-2001
DEFINITION Homo sapiens genomic DNA, chromosome 6q25.2, clone:KB1F5.
ACCESSION AP001576
VERSION AP001576.3 GI:15208253
KEYWORDS
SOURCE Homo sapiens cell_line:FLNB 14 - 14 DNA, clone_lib:keio BAC library
clone:KB1F5.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 105199)
AUTHORS Shimizu,N. and Asakawa,S.
TITLE Homo sapiens DNA chromosome 6 SEQUENCE
JOURNALS Published Only in Database (2000) In press
AUTHORS Shimizu,N. and Asakawa,S.
REFERENCE 2 (bases 1 to 105199)
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
On Aug 16, 2001 this sequence version replaced gi:8096256.
COMMENT
FEATURES
location/Qualifiers
source 1..105199
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q25.2"
/clone="KB1F5"
/cell_line="FLNB 14 - 14"
/clone_lib="Keio BAC library"
166..474
/rpt_family="Alusx"
/evidence=not_experimental
520..590
/rpt_family="AT-rich"
/evidence=not_experimental
complement(912..1017)
/rpt_family="FLAN_C"
/evidence=not_experimental
1199..1490
/rpt_family="MLT1F"
/evidence=not_experimental
1844..1922
/rpt_family="(TATATG)n"

```

repeat_region	/evidence-not_experimental 2732. .3028 /rpt_family="AluSx"	repeat_region	18854. .18955 /rpt_family="AT_rich"
repeat_region	/evidence-not_experimental complement(3523. .3404) /rpt_family="L2"	repeat_region	/evidence-not_experimental 19314. .19624 /rpt_family="AluSx"
repeat_region	/evidence-not_experimental complement(3608. .3816) /rpt_family="L2"	repeat_region	/evidence-not_experimental 20973. .21071 /rpt_family="MTR"
repeat_region	/evidence-not_experimental complement(4573. .4720) /rpt_family="MIR"	repeat_region	/evidence-not_experimental complement(22061. .22369) /rpt_family="AluSx"
repeat_region	/evidence-not_experimental complement(7641. .8105) /rpt_family="L1PA16"	repeat_region	/evidence-not_experimental 22428. .22463 /rpt_family="AT_rich"
repeat_region	/evidence-not_experimental 8106. .8852 /rpt_family="L1PA16"	repeat_region	/evidence-not_experimental complement(22469. .22781) /rpt_family="AluSx"
repeat_region	/evidence-not_experimental 8855. .8884 /rpt_family="AT_rich"	repeat_region	/evidence-not_experimental complement(22820. .23154) /rpt_family="MER7A"
repeat_region	/evidence-not_experimental 9163. .9192 /rpt_family="AT_rich"	repeat_region	/evidence-not_experimental 23362. .23397 /rpt_family="AT_rich"
repeat_region	/evidence-not_experimental 9351. .9656 /rpt_family="AluSx"	repeat_region	/evidence-not_experimental 23468. .23503 /rpt_family="AT_rich"
repeat_region	/evidence-not_experimental complement(10418. .10545) /rpt_family="MER58C"	repeat_region	/evidence-not_experimental 24148. .24175 /rpt_family="(A)n"
repeat_region	/evidence-not_experimental 10546. .10585 /rpt_family="Alu"	repeat_region	/evidence-not_experimental complement(24576. .24879) /rpt_family="AluSx"
repeat_region	/evidence-not_experimental 10708. .10797 /rpt_family="L2"	repeat_region	/evidence-not_experimental 24980. .25344 /rpt_family="MLT2E"
repeat_region	/evidence-not_experimental complement(10837. .11091) /rpt_family="AluSc"	repeat_region	/evidence-not_experimental complement(25365. .25422) /rpt_family="MTR"
repeat_region	/evidence-not_experimental 11771. .11813 /rpt_family="AT_rich"	repeat_region	/evidence-not_experimental complement(25372. .25487) /rpt_family="L2"
repeat_region	/evidence-not_experimental 12272. .12292 /rpt_family="AT_rich"	repeat_region	/evidence-not_experimental 25515. .25637 /rpt_family="MER9A"
repeat_region	/evidence-not_experimental complement(12629. .12706) /rpt_family="MTR"	repeat_region	/evidence-not_experimental 25822. .25868 /rpt_family="Alu"
repeat_region	/evidence-not_experimental 12903. .13167 /rpt_family="AluJo"	repeat_region	/evidence-not_experimental 26514. .26818 /rpt_family="AluYa5"
repeat_region	/evidence-not_experimental 13173. .13453 /rpt_family="AluSg"	repeat_region	/evidence-not_experimental 27318. .27344 /rpt_family="(TTTG)n"
repeat_region	/evidence-not_experimental 13871. .13920 /rpt_family="AT_rich"	repeat_region	/evidence-not_experimental 27983. .28027 /rpt_family="(TTTG)n"
repeat_region	/evidence-not_experimental complement(14938. .15193) /rpt_family="AluSg"	repeat_region	/evidence-not_experimental complement(28187. .28382) /rpt_family="MER20"
repeat_region	/evidence-not_experimental 15379. .15453 /rpt_family="L2"	repeat_region	/evidence-not_experimental 28857. .28993 /rpt_family="FLAM_C"
repeat_region	/evidence-not_experimental complement(15501. .15754) /rpt_family="LTR37B"	repeat_region	/evidence-not_experimental 29511. .29565 /rpt_family="(TG)n"
repeat_region	/evidence-not_experimental complement(16177. .16924) /rpt_family="MER31-internal"	repeat_region	/evidence-not_experimental 29582. .29620 /rpt_family="AT_rich"
repeat_region	/evidence-not_experimental complement(17232. .17377) /rpt_family="MER31-internal"	repeat_region	/evidence-not_experimental 29781. .29923 /rpt_family="MTR"
repeat_region	/evidence-not_experimental	repeat_region	/evidence-not_experimental complement(30500. .30659)

```
/rpl_family="MIR"
/evidence-not_experimental
complement(30851..31133)
/rpl_family="THE1C"
/evidence-not_experimental
31306..31348
/rpl_family="AT-rich"
/evidence-not_experimental
32634..32725

repeat_region
repeat_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
us-09-697-089-2 x AP001576/rev ..

Align seg 1/1 to reverse of: AP001576 from: 1 to: 105199

64
|||||
42661 GAGTCTGTAACCTTGTTCTTAAAGC 42635

seq_documentation_block:
LOCUS      HSDJ93K22 113804 bp      DNA      PRI      17-MAR-2000
DEFINITION Human DNA sequence from clone RPI-93K22 on chromosome 6q14.1-15.3.
            Contains the gene for a novel protein (contains DKFZ564B116),
            ESTs, STSs, GSSs and a putative Cpg island, complete sequence.
ACCESSION  AL050333.18 GI:6911641
VERSION    AL050333.18 GI:6911641
KEYWORDS   HTG; Cpg island.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 113804)
AUTHORS   Tracey, A.
TITLE      Direct Submission
JOURNAL    Submitted (15-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequests@sanger.ac.uk
            On Feb 7, 2000 this sequence version replaced gi:6807622.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WormPep; Information
            on the WormPep database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            Rpi-93K22 is from the library RPI-1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RPI-93K22 The true left
            end of clone RPI-492P14 is at 84788 in this sequence.
            Location/Qualifiers

source
1..113804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q14.1-15.3"
/clone="RPI-93K22"
/clone_11b="RPI-1"
691..955
/note="12 repeat: matches 1733..2019 of consensus"
1289..1574
/note="Alu1b repeat: matches 2..289 of consensus"
2115..2429
/note="Alu1b repeat: matches 1..312 of consensus"
4453..4811
/note="match: STS: Em:267262"
4645..4698
/note="27 copies 2 mer ac 96% conserved"
7184..7560
/note="MER47B repeat: matches 1..370 of consensus"
7576..7710
/note="Alu1b repeat: matches 163..300 of consensus"
8457..8552
/note="L1m1 repeat: matches 6113..6224 of consensus"
8570..8866
/note="Alu1 repeat: matches 1..294 of consensus"
8915..10626
/note="L1p5 repeat: matches 3340..5025 of consensus"
10637..11821
/note="L1p2 repeat: matches 11..1190 of consensus"
11822..12127
/note="Alu1 repeat: matches 1..307 of consensus"
12128..17094
/note="L1p2 repeat: matches 1190..6144 of consensus"
17098..17194
/note="L1p5 repeat: matches 5002..5101 of consensus"
17196..17328
/note="L1m4A repeat: matches 5916..6044 of consensus"
17326..18325
/note="L1p5 repeat: matches 5116..6060 of consensus"
18364..18841
/note="L1m4A repeat: matches 5813..6294 of consensus"
19026..19172
/note="MIR repeat: matches 97..260 of consensus"
19699..19740
/note="21 copies 2 mer gt 97% conserved"
20261..20639
/note="THE1C repeat: matches 1..371 of consensus"
complement(20948..21434)
/note="match: GSS: Em:A0777802"
22356..22656
/note="Alu1 repeat: matches 1..298 of consensus"
22575..22991
/note="match: GSS: Em:AQ013057"
23303..23611
/note="Alu1 repeat: matches 1..308 of consensus"
23792..24296
/note="match: GSS: Em:AQ129613"
23853..24132
/note="HAL1 repeat: matches 775..1074 of consensus"
24137..24427
/note="Alu1 repeat: matches 1..302 of consensus"
25324..25471
/note="MER5A repeat: matches 18..176 of consensus"
25788..26095
/note="Alu1b repeat: matches 1..297 of consensus"
26138..26373
/note="L1R33 repeat: matches 253..513 of consensus"
26519..26549
/note="MER5A repeat: matches 146..176 of consensus"
26799..27850
/note="L1m4 repeat: matches 2682..3765 of consensus"
28354..28642
/note="Alu1x repeat: matches 1..291 of consensus"
```

```

repeat_region
29048..29145
/note="Alu repeat: matches 204..299 of consensus"
complement(join(29950..31268,32092..32224,33079..33150,
41591..41740,50785..50928,51499..51591,54191..54354,
56010..56158,59853..59965,60299..60420,61134..61194,
62240..62357,64569..64720,72419..72505,73939..74541,
75788..75962,77672..77849,80302..80425,83152..83332,
83780..83897,85189..85359,86904..87014,91430..91554,
93871..93967,99878..10055,107274..107312))
/gene="d93k2.1"
/note="match: CDNAs: Em:AL050018; match: ESTs: Em:AL042768
Em:AA258921 Em:HSC19C091 Em:R13393 Em:AA242432 Em:AA746291
Em:T32527 Em:AA428674 Em:R59046 Em:HSC2RD011 Em:HSB45H042
Em:Z19312 Em:AI556384 Em:T11904 Em:R1445866 Em:T11903
Em:R148858 Em:AA258158 Em:AA971694 Em:HSC19C092 Em:AI827319
Em:AA861452 Em:AA007318 Em:AA403171 Em:AA401235
Em:AA403173 Em:R17235 Em:AI393859 Em:AA406086 Em:AA428185
Em:AI540581 Em:AI382812 Em:AI952901 Em:AI816830
Em:AI277356 Em:AA461143 Em:H06057 Em:T50258 Em:AM043840
Em:HA2659 Em:R41845 Em:AA236615 Em:R1140662 Em:HSC2RD012
Em:HSB13C042 Em:AI458423 Em:R1140997 Em:AA639828 Em:H92102
Em:AM023058 Em:AI669354 Em:AI525340 Em:AI608787 Em:H21698
Em:AI469859"
/evidence=not_experimental
/product="d93k2.1 (novel protein (contains
DKF2P564B16))"
complement(29950..107312)
/gene="d93k2.1"
complement(29950)
/gene="d93k2.1"
29967..30202
/note="match: SRS: Em:T50208"
29968..30202
/note="match: SRS: Em:G11689"
complement(29971..29976)
/gene="d93k2.1"
complement(join(31137..31268,32092..32224,33079..33150,
41591..41740,50785..50928,51499..51591,54191..54354,
56010..56158,59853..59965,60299..60420,61134..61194,
62240..62357,64569..64720,72419..72505,73939..74541,
75788..75962,77672..77849,80302..80425,83152..83332,
83780..83897,85189..85359,86904..87014,91430..91554,
93871..93967,99878..100186))
/gene="d93k2.1"
/note="exon 12 based purely on GENESH and GENSCAN
prediction; match: proteins: Tr:Q913T8"
/codon_start=1
/evidence=not_experimental
/product="d93k2.1 (novel protein (contains
DKF2P564B16))"
/protein_id="CAB81625.1"
/db_xref="GI:7263990"
/translacion="MPDCTSKGRSLKHALDLVSVTKGSENOIKAFLSHCYNATIK
DVEGRNALHLVSSCGKGVLDWLIOKGVLDLVKDKSGWTALHRSIFYGHIDCVSL
KIGVSLYIDDKGLSALDLYMKDRPHVYKMTDPTDYTWGDNFTTGHSSONSIL
HBEIYDLFSRSGIYIKOVYLCFHSYELSKOGVYTCGPGRLGGEORCLNVRK
VGLNGHNCVAAAKDHVYLTEDGCYVTRGLNTHPHOGLTIPRPSCAVROIOAKY
LKGRITICVAAGRFHYLTREAVYTMGNGCOLGLDPRNKECYTAROVYSAHLHK
DIALSLVAASDAVTCVTRGDIYTLADYOCKRMASKQNLNKKVLVSGHMEKYDPE
HLKENGCKTICLAMDAGAVFCRWRSVNSLQCRWAYAPROVEISDIALNREILFET
ODGEGFRGRMFEERKRSERKEILSNLHNSSDVSYSDINSYERIRLEKLEIFARA
VASVTDPSGNCNALIOSDKPTSLYEIPAVSSSFEFEKGLREADEMSHIDVFOV
GNRLPAHKYIILVHSDPFROKTLSDGNTSEPTDIYOKRDLKSGCLPVEVYHPDMF
EYLLOPIYTDCTDELTHGFRPHIHLKNPEEYOGTINLKNVNRHEDNOQSAEYV
KSNQOYTVSEROKSKPKCKKRNIREDPVMLQVAKKEPNSLSSLDKRENE
KINVLAKNTGNLKLQSKRWVYKEKVMLEFLESONVPEISVLVAQDLITR
LKEICVALLTEKTLKNAAMLLEFAMYSAKOLKSCLOFGLNMAALLAASLDVLS
DKVLDLSEYRKMIIPAMDRVITYPDGDPISTYLEVEDGDIFLKEIMEHNSHSTM
FKAKTKAKKRRKSDSGYNLSDILOSPSTGILKGNKNSVSLPELITSDSEG
SYAGNSPRDLOSPOPTTFPHSKDIKAKYKPYVNGSPYSSKEDLKPMKPSPLIKSA
POPISNRIIDTSSASWAGSFPVSPVADVRLTIELESKCCATPKSHDGKTVS
HGVKLQOKRKMIALTKEKNSGMSMEVLETPSPAPKPVAAVMASSLSVSSKSRD
FLLEKKSVTSHSGDVKVSKFGIENSQAKIYKCSYHGTGPGEGNHTSIDPLIDS

```

```

misc_feature
complement(join(32151..32224,33079..33150,41591..41745))
/gene="d93k2.1"
/note="match: SRS: Em:219312"
32708..33013
/note="Alus repeat: matches 1..306 of consensus"
complement(32966..33523)

repeat_region
repeat_feature
misc_feature
alignment_scores:
  Quality: 9.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Percent Identity: 100.000
alignment_block:
  US-09-697-089-2 x HSDJ93K22 ..
Align seg 1/1 to: HSDJ93K22 from: 1 to: 113804
703 SerLeuSerLeuValIleuSerThrcys 711
|||||
87507 TCCTTATCTTGGTTTATCAACATGT 87533

seq_name: gb-pr:HS274L7
seq_documentation_block:
LOCUS HS274L7 119566 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from PAC 274L7 on chromosome X contains ESTs.
ACCESSION 282195
VERSION 282195.1 GI:1841909
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 119566)
REFERENCE
AUTHORS Mistry, S.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Feb 17, 1997 this sequence version replaced gi:1772944.
de Jong P.J., enquiries: http://bacpac.med.bufileo.edu/IMPORTANT:
This sequence is the entire insert of clone 274L7. This sequence
has been finished according to sequence map criteria as follows.
An attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 274L7 is at 1 in this sequence. The true
right end of clone 274L7 is at 119566.
274L7 is from the human PAC library described in Ioannou A.P. et al
Nature Genet 6, 84-89.
FEATURES
Location/Qualifiers
source 1..119566
organism="Homo sapiens"
db_xref="taxon:9606"
chromosome="X"
map="X"
/clone_lib="RPCT-1"
/clone="RP1-274L7"
1..266
/note="MULTI2 repeat: matches 112..374 of consensus"
333..539
/note="MIR repeat: matches 75..251 of consensus"
635..755
/note="MIR repeat: matches 28..148 of consensus"
851..919
/note="MULT2CA repeat: matches 71..1 of consensus"
1150..1260
/note="MIR repeat: matches 47..157 of consensus"
2295..2338
repeat_region

```

```
repeat_region /note="MLT2E repeat: matches 397. .349 of consensus"
2432. .2722
/note="MLT2FB repeat: matches 366. .79 of consensus"
repeat_region 2723. .3016
/note="Alusq repeat: matches 1. .301 of consensus"
repeat_region 3017. .3112
/note="MLT2G repeat: matches 97. .1 of consensus"
4422. .4730
/note="MLT2D repeat: matches 5. .319 of consensus"
4675. .4787
/note="MLT2CB repeat: matches 307. .421 of consensus"
repeat_region 5319. .5392
/note="MLT2G repeat: matches 316. .388 of consensus"
5323. .5328
/note="single clone"
unsure 5353. .5392
/note="MLT2E repeat: matches 358. .397 of consensus"
5570. .6048
/note="MLT1C repeat: matches 466. .1 of consensus"
repeat_region 6251. .6298
/note="MLT1C repeat: matches 96. .144 of consensus"
6897. .7124
/note="MLT1 repeat: matches 262. .17 of consensus"
8553. .8726
/note="Aluio repeat: matches 302. .124 of consensus;
incomplete repeat"
repeat_region 8762. .8856
/note="MLT1 repeat: matches 247. .154 of consensus"
8857. .9153
/note="Alusq repeat: matches 303. .1 of consensus"
9155. .9254
/note="MLT1 repeat: matches 158. .58 of consensus"
12658. .12835
/note="MLT1 repeat: matches 2. .191 of consensus"
12835. .13050
/note="MLT1 repeat: matches 261. .49 of consensus"
13310. .13422
/note="FRAM repeat: matches 139. .21 of consensus"
14239. .14349
/note="MLT1 repeat: matches 260. .147 of consensus"
17476. .17602
/note="MLT2 repeat: matches 25. .146 of consensus"
17756. .18071
/note="Alusq repeat: matches 1. .300 of consensus"
18093. .18177
/note="MLT1 repeat: matches 72. .154 of consensus"
18567. .18700
/note="BC200 repeat: matches 1. .136 of consensus"
20457. .20945
/note="MLT1C repeat: matches 1. .466 of consensus"
21470. .21609
/note="MLT1 repeat: matches 96. .262 of consensus"
22097. .22394
/note="MLT1 repeat: matches 1. .302 of consensus"
22763. .22840
/note="Aluio repeat: matches 296. .223 of consensus;
incomplete repeat"
22841. .23024
/note="Aluio repeat: matches 184. .2 of consensus;
incomplete repeat"
23402. .23682
/note="MER42C repeat: matches 1538. .1273 of consensus"
24163. .24247
/note="MLT1B repeat: matches 16. .102 of consensus"
24230. .24431
/note="MLT1C repeat: matches 159. .353 of consensus"
29569. .29857
/note="Alusq repeat: matches 3. .302 of consensus"
29942. .30067
/note="MLT1C repeat: matches 338. .464 of consensus"
30070. .30249
/note="MLT1-INTERNAL repeat: matches 1. .179 of consensus"
30192. .30548
repeat_region /note="MLT1-INTERNAL repeat: matches 239. .609 of
consensus"
30558. .30861
/note="Aluub repeat: matches 1. .302 of consensus"
30872. .31157
/note="MLT1-INTERNAL repeat: matches 614. .885 of
consensus"
repeat_region 31197. .34221
/note="G rich region 36% G, 61% G+C"
32354. .32468
/note="single clone"
unsure 34334. .34641
/note="Alusq repeat: matches 300. .1 of consensus"
34645. .34916
/note="MLT1-INTERNAL repeat: matches 1056. .1337 of
consensus"
34925. .35393
/note="MLT1C repeat: matches 1. .466 of consensus"
36316. .36617
/note="Aluv repeat: matches 301. .1 of consensus"
36618. .36941
/note="L1 repeat: matches 2410. .2081 of consensus"
37211. .37407
/note="L1PA15 repeat: matches 707. .904 of consensus"
38123. .38217
/note="MLT2 repeat: matches 60. .144 of consensus"
39321. .39379
/note="MADE1 repeat: matches 1. .75 of consensus"
39598. .40490
/note="L1PA2 repeat: matches 893. .1 of consensus"
40341. .44844
/note="L1 repeat: matches 5390. .896 of consensus"
44840. .45617
/note="L1 repeat: matches 772. .4 of consensus"
47899. .48199
/note="Alusq repeat: matches 302. .2 of consensus"
49167. .49468
/note="Aluub repeat: matches 302. .1 of consensus"
49577. .50391
/note="MER11A repeat: matches 4. .847 of consensus"
49953. .50583
/note="MER11B repeat: matches 1. .633 of consensus"
50783. .51074
/note="Alusp repeat: matches 293. .1 of consensus"
52134. .52440
/note="Aluio repeat: matches 302. .1 of consensus"
53637. .53936
/note="Alusq repeat: matches 297. .1 of consensus"
54617. .54673
/note="Alusp repeat: matches 57. .1 of consensus;
incomplete repeat"
55958. .55993
/note="Alusq repeat: matches 1. .36 of consensus;
incomplete repeat"
55994. .56058
/note="Aluio repeat: matches 230. .294 of consensus;
incomplete repeat"
56681. .56718
/note="Aluv repeat: matches 38. .1 of consensus; incomplete
repeat"
56897. .57197
/note="Alusq repeat: matches 1. .301 of consensus"
57440. .57711
/note="Alusq repeat: matches 5. .300 of consensus"
59123. .59426
/note="Aluv repeat: matches 1. .301 of consensus"
59975. .60219
/note="Aluio repeat: matches 49. .300 of consensus;
incomplete repeat"
61656. .61955
/note="Aluub repeat: matches 1. .301 of consensus"
62288. .62580
/note="Alusq repeat: matches 17. .303 of consensus;
```

repeat_region 63146..63447 Incomplete repeat"
/note="AluSg repeat: matches 1. .289 of consensus"
repeat_region 64381..64677
/note="AluSx repeat: matches 8. .299 of consensus"
prim_transcript complement(65431..65493)
/note="match: 3' EST C00574"

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

align seg 1/1 to: HSJ274L7 from: 1 to: 119566
US-09-697-089-2 x HSJ274L7

700 ValAlaGlySerLeuSerLeuValLeu 708
|||||
65907 GTTCCTGAGACCTTCTCTGCTTTTG 65933

seq_name: gb_htg:HSJ282H10

seq_documentation_block:
LOCUS HSJ282H10 120029 bp DNA HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 6 clone RP1-282H10 map p22.1-22.3, ***
SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION AL132672
VERSION AL132672.14 GI:14348905
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 120029)
Garner, P.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On Jun 12, 2001 this sequence version replaced gi:12331282.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: dj282H10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 0% of reads
Sequencing vector: plasmid; L08752; 98% of reads
Chemistry: Dye-terminator ABI; 1% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Consensus quality: 118224 bases at least Q40
Consensus quality: 118449 bases at least Q30
Insert size: 119629; sum-of-contigs
Insert size: 117431; 9.3% error; agarose-fp
Quality coverage: 18.41x in Q20 bases; sum-of-contigs Quality
coverage: 19.36x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 23429: contig of 23429 bp in length
* 23430 23529: gap of 100 bp

23530 63600: contig of 40071 bp in length
* 63601 63700: gap of 100 bp
* 63701 100099: contig of 36399 bp in length
* 100100 100199: gap of 100 bp
* 100200 102366: contig of 2167 bp in length
* 102367 102466: gap of 100 bp
* 102467 120029: contig of 17563 bp in length.
Location/Qualifiers
1..120029
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p22.1-22.3"
/clone="RP1-282H10"
/clone_1fb="RP1-1"
1..23429
/note="assembly_fragment:05837
clone_end:SP6
vector_side:left"
23530..63600
/note="assembly_fragment:00423
fragment_chain:1"
63701..100099
/note="assembly_fragment:02685
fragment_chain:1"
100200..102366
/note="assembly_fragment:01789"
102467..120029
/note="assembly_fragment:02341
clone_end:T7
vector_side:right"

BASE COUNT 36483 a 24840 c 24637 g 33666 t 403 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

align seg 1/1 to: HSJ282H10 from: 1 to: 120029
US-09-697-089-2 x HSJ282H10

451 AlAGlyArgArgLeuSerLeuValLeu 459
|||||
21903 GCTGGCCGCCGCTCAGCTGCTCTT 21929

seq_name: gb_pr:AL359833

seq_documentation_block:
LOCUS AL359833 122014 bp DNA PRI 23-NOV-2000
DEFINITION Human DNA sequence from clone RP11-186C9 on chromosome 1, complete
sequence.
ACCESSION AL359833
VERSION AL359833.12 GI:11340294
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 122014)
Hall, R.
Direct Submission
Submitted (23-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On Nov 24, 2000 this sequence version replaced gi:11225780.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrl>

RP11-186C9 is from the library RPCT-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://Dacpac.med.buffalo.edu/>

VECTOR: PBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-186C9 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-186C9 is at 122014 in this sequence. The true left end of clone RP11-421110 is at 754 in this sequence. The true right end of clone RP11-453022 is at 120 in this sequence.

FEATURES

SOURCE

```
1. 122014
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-186C9"
/clone_id="RPCT-11.1"
122..461
/note="match: GSS: Em:A0276589"
877..1115
/note="match: GSS: Em:A0553244"
1644..2026
/note="LMC repeat: matches 1499..1894 of consensus"
1970..2471
/note="match: GSS: Em:A0559088"
2303..2675
/note="MLT1B repeat: matches 1..390 of consensus"
2907..3584
/note="LMC repeat: matches 2386..2728 of consensus"
3697..4050
/note="LIM4 repeat: matches 2956..3318 of consensus"
4223..4310
/note="LIM4 repeat: matches 4840..4930 of consensus"
4311..4621
/note="AluX repeat: matches 1..312 of consensus"
4623..4662
/note="10 copies 4 mer gaaa 80% conserved"
5403..5662
/note="AluX repeat: matches 1..276 of consensus"
6453..6706
/note="MIR repeat: matches 8..262 of consensus"
complement(6516..6989)
/note="match: GSS: Em:A0798812"
8532..8623
/note="L2 repeat: matches 2088..2184 of consensus"
complement(8619..9104)
/note="match: GSS: Em:A0242845"
8771..9054
/note="AluX repeat: matches 1..280 of consensus"
9394..9911
/note="MLT1D repeat: matches 1..505 of consensus"
10665..10733
/note="L2 repeat: matches 2673..2749 of consensus"
10810..11093
/note="AluX repeat: matches 1..284 of consensus"
```

```
repeat_region
11275..11491
/note="L2 repeat: matches 2453..2678 of consensus"
11492..11812
/note="AluY repeat: matches 1..311 of consensus"
11813..12079
/note="AluJo repeat: matches 34..307 of consensus"
12080..12295
/note="L2 repeat: matches 2263..2453 of consensus"
13150..13312
/note="MIR repeat: matches 26..206 of consensus"
14843..15369
/note="match: STS: Em:G59154
match: GSS: Em:A0347511"
17082..17117
/note="L2 copies 3 mer atg 86% conserved"
17125..17345
/note="MIR repeat: matches 8..256 of consensus"
complement(17254..17781)
/note="match: GSS: Em:A0712657"
17575..17763
/note="MER53 repeat: matches 2..189 of consensus"
18286..18366
/note="L2 repeat: matches 2629..2701 of consensus"
18338..18454
/note="L2 repeat: matches 2580..2710 of consensus"
18576..18725
/note="MIR repeat: matches 41..204 of consensus"
18922..19041
/note="MIR repeat: matches 1..120 of consensus"
19042..19353
/note="AluSg repeat: matches 1..309 of consensus"
19354..19486
/note="MIR repeat: matches 120..252 of consensus"
20467..20574
/note="27 copies 4 mer aaag 70% conserved"
21001..21096
/note="L2 repeat: matches 2416..2502 of consensus"
21398..21503
/note="MIR repeat: matches 21..132 of consensus"
21919..21981
/note="LIM43 repeat: matches 5890..5951 of consensus"
22031..22133
/note="L2 repeat: matches 2588..2709 of consensus"
22311..22344
/note="17 copies 2 mer tt 88% conserved"
25897..25952
/note="14 copies 4 mer gtgt 91% conserved"
26855..26958
/note="MIR repeat: matches 2..96 of consensus"
27193..27478
/note="AluJo repeat: matches 1..282 of consensus"
27482..27618
/note="MIR repeat: matches 66..195 of consensus"
complement(27724..28165)
/note="match: GSS: Em:A0680865"
complement(27740..28155)
/note="match: GSS: Em:A0681916"
30489..30554
/note="33 copies 2 mer aa 66% conserved"
31186..31219
/note="17 copies 2 mer gt 85% conserved"
31188..31219
/note="8 copies 4 mer ggt 87% conserved"
32209..32257
/note="L2 repeat: matches 2702..2750 of consensus"
32281..32339
/note="MIR repeat: matches 110..168 of consensus"
32367..32477
/note="MIR repeat: matches 148..262 of consensus"
34251..34437
/note="MERSA repeat: matches 1..189 of consensus"
34326..34719
/note="match: GSS: Em:A0186533"
```



```

repeat_region      35098. .35173
                    /note="MIR repeat: matches 59. .145 of consensus"
repeat_region      35418. .35735
                    /note="AluBq repeat: matches 4. .311 of consensus"
repeat_region      36438. .36483
                    /note="23 copies 2 mer ga 80% conserved"
repeat_region      36440. .36483
                    /note="11 copies 4 mer gaga 81% conserved"
repeat_region      36691. .37152
                    /note="MLTIF repeat: matches 1. .466 of consensus"
repeat_region      37053. .37465
                    /note="AluDb repeat: matches 1. .295 of consensus"
repeat_region      37702. .37869
                    /note="MLTIF repeat: matches 56. .233 of consensus"
repeat_region      37926. .38145
                    /note="MLTIF repeat: matches 232. .460 of consensus"
misc_feature        complement(38267. .38728)
                    /note="match: GSS: Em:A0123310"
repeat_region      38432. .38480
                    /note="MLTIF repeat: matches 485. .533 of consensus"
misc_feature        38734. .39111
                    /note="match: GSS: Em:A0278876"
repeat_region      39817. .39968
                    /note="L2 repeat: matches 2591. .2750 of consensus"
repeat_region      40082. .40232
                    /note="L2 repeat: matches 2310. .2470 of consensus"
repeat_region      41093. .41206
                    /note="38 copies 3 mer gga 90% conserved"
repeat_region      41871. .42064
                    /note="L1M4 repeat: matches 5148. .5344 of consensus"
repeat_region      42275. .42694
                    /note="L1M4 repeat: matches 3812. .4238 of consensus"
repeat_region      43257. .43288

```

```
alignment_scores:      .
  quality:      9.00      Length:      9
  ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.0000
```

alignment_block:

US-09-697-089-2 x AL359833/rev

Align seg 1/1 to reverse of: AL359833 from: 1 to: 122014

326 GlnLysSerArgCysLeuArgAsnLeu 334

60936 CAAAGTCCAGATGTTGAGAAACCTC 60910

seq_name: gb_pr:AC003085

```
seq_documentation_block:
```

DEFINITION Human BAC clone CTB-94H21 from 7q21-q22, complete sequence.

VERSION AC003085.1 GI:2588620

SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 124836)

TITLE The sequence of H. sapiens BAC clone CTB-94H21

REFERENCE 2 (bases 1 to 124836)

TITLE

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

AUTHORS
WITH

JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington

REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS	4 (bases 1 to 124836)
TITLE	Waterston, R.
JOURNAL	Direct Submission
COMMENT	Submitted (04-FEB-2000) Department of Genetics, Washington
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
	----- genome center

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRV/CHR7> or send mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
clone CTR-94H21 is from the first release of the human BAC library
CTTB-978SK-B. The library contains cloned DNA from the male
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.
Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
(1996). This clone is available from Research Genetics, Inc.
(http://www.resgen.com).
VECTOR: pBelOAC11
selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of CTB-94H21;
the actual end is at base position 124836 of CTB-94H21.

This clone contains STS SWSS1530 (NID:g485245) and SWSS784 (NID:g485556).

FEATURES	Location/Qualifiers
source	1. .124836

```
repeat_region      complement(1..798)
                    /rpt_family="HHR"
repeat_region      complement(804..1156)
                    /rpt_family="L1"
repeat_region      complement(1157..1508)
                    /rpt_family="MER"
repeat_region      complement(4115..4133)
                    /rpt_family="L1"
repeat_region      complement(4154..4445)
                    /rpt_family="ALU"
repeat_region      complement(5099..5495)
```

```

repeat_region /rpl_family="L1"
6346..6635
/rpl_family="ALU"
complement(7195)
repeat_region /rpl_family="L1"
complement(10424)
repeat_region /rpl_family="L1"
complement(11139)
repeat_region /rpl_family="L1"
complement(11161)
repeat_region /rpl_family="ALU"
complement(11655)
repeat_region /rpl_family="L1"
complement(12595)
repeat_region /rpl_family="ALU"
complement(15938)
repeat_region /rpl_family="L1"
complement(16398)
repeat_region /rpl_family="L1"
complement(24943)
repeat_region /rpl_family="ALU"
complement(25660)
repeat_region /rpl_family="L1"
26634..27052
/rpl_family="L1"
27212..27239
/rpl_family="L1"
complement(34045)
repeat_region /rpl_family="L1"
34136..34157
/rpl_family="L1"
complement(34201)
repeat_region /rpl_family="L1"
36116..36213
/rpl_family="ALU"
complement(37632)
repeat_region /rpl_family="ALU"
40491..40797
/rpl_family="ALU"
complement(41017)
repeat_region /rpl_family="MER"
complement(41263)
repeat_region /rpl_family="ALU"
complement(41555)
repeat_region /rpl_family="MER"
complement(44548)
repeat_region /rpl_family="ALU"
complement(45173)
repeat_region /rpl_family="ALU"
46832..47258
/rpl_family="ALU"
complement(47844)
repeat_region /rpl_family="ALU"
50858..50894
/rpl_family="L1"
52434..52581
/rpl_family="L1"
52583..52868
/rpl_family="ALU"
52880..53388
/rpl_family="L1"
53812..55018
/rpl_family="L1"
55039..55692
/rpl_family="L1"
complement(57221)
repeat_region /rpl_family="L1"
58274..58361
/rpl_family="L1"
58363..58654
/rpl_family="ALU"

```

```

repeat_region 58655..58850
/rpl_family="L1"
complement(58813)
repeat_region /rpl_family="L1"
60783..61071
/rpl_family="ALU"
complement(61076)
repeat_region /rpl_family="L1"
complement(65373)
repeat_region /rpl_family="L1"
65400..66294
/rpl_family="L1"
66378..66452
/rpl_family="L1"
67627..67918
/rpl_family="ALU"
67919..67963
/rpl_family="L1"
69743..70032
/rpl_family="ALU"
70043..71220
/rpl_family="L1"
complement(73454)
repeat_region /rpl_family="L1"
73867..74124
/rpl_family="ALU"
74483..74771
/rpl_family="ALU"
complement(75034)
repeat_region /rpl_family="L1"
76157..76446
/rpl_family="ALU"

```

```

alignment_scores:
  Quality: 9.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Percent Identity: 100.000

```

```

alignment_block:
  US-09-697-089-2 x AC003085

```

```

Align seg 1/1 to: AC003085 from: 1 to: 124836

```

```

919 LeuThraspThrcGlutleArqyleu 927
|||||
17004 CTRACGTGATACGAATAAGATTTTA 17030

```

```

seq_name: gb_htg:AC015608

```

```

seq_documentation_block:
  LOCUS AC015608 132466 bp DNA HTG 07-JUN-2000
  DEFINITION Homo sapiens clone RPL1-45F5, WORKING DRAFT SEQUENCE, 11 unordered
  pieces.
  ACCESSION AC015608
  KEYWORDS AC015608.5 GI:8315560
  SOURCE HTG: HTGS_PHASE1; HTGS_DRAFT.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
  1 (bases 1 to 132466)
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
  Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckgatter,B.,
  Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
  Cooke,P., Dearfllano,K., Dewar,K., Domino,M., Deland,L., Doyle,M.,
  Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
  Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
  Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
  Lehoccky,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,

```

TITLE
JOURNAL

COMMENT

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J., Testaye, S., Tirrell, A., Vassilev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced gi:7657800.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----
Center project name: L1140
Center clone name: 45.F.5

----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 125885 bases at least Q40
Consensus quality: 128872 bases at least Q30
Consensus quality: 129978 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 131466; sum-of-ctrls
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 6.0 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1271: contig of 1271 bp in length
* 1272 1371: gap of 100 bp
* 1372 2569: contig of 1198 bp in length
* 2570 2669: gap of 100 bp
* 2670 4080: contig of 1411 bp in length
* 4081 4180: gap of 100 bp
* 4181 5304: contig of 1124 bp in length
* 5305 5404: gap of 100 bp
* 5405 8687: contig of 3283 bp in length
* 8688 8787: gap of 100 bp
* 8788 14434: contig of 5647 bp in length
* 14435 14534: gap of 100 bp
* 14535 18607: contig of 4073 bp in length
* 18608 18707: gap of 100 bp
* 18708 35116: contig of 16409 bp in length
* 35117 35216: gap of 100 bp
* 35217 51905: contig of 16689 bp in length
* 51906 52005: gap of 100 bp
* 52006 87674: contig of 35669 bp in length
* 87675 87774: gap of 100 bp
* 87775 132466: contig of 44692 bp in length.

FEATURES
source
1. 132466
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-45F5"
/clone_lib="RPC1-11 Human Male BAC"
1. 1271
/note="assembly-fragment"
misc_feature 1372..2569
/note="assembly-fragment"
misc_feature 2670..4080

misc_feature /note="assembly-fragment"
4181..5304
/note="assembly-fragment"
misc_feature 5405..8687
/note="assembly-fragment"
misc_feature 8788..14434
/note="assembly-fragment"
misc_feature 14535..18607
/note="assembly-fragment"
misc_feature 35217..51905
/note="assembly-fragment"
misc_feature 52006..87674
/note="assembly-fragment"
misc_feature 87775..132466
/note="assembly-fragment"

BASE COUNT 40541 a 24980 c 24933 g 41008 t 1004 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AC015608/rev ..

Align seg 1/1 to reverse of: AC015608 from: 1 to: 132466

326 GlnysSerArgCysLeuArgAsnLeu 334
|||||||
55181 CAAAGTCAGAGTGTTCGAGAACCTC 55155

seq_name: gb_hhg:AC010625

seq_documentation_block:
LOCUS AC010625 138638 bp DNA HTG 19-APR-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2128F4, WORKING DRAFT SEQUENCE,
4 ordered pieces.
ACCESSION AC010625 GI:9256261
VERSION AC010625.4 GI:9256261
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 138638)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 138638)
DOE Joint Genome Institute.
TITLE Direct Submission
SUBMITTED (16-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710611.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 676825
Center clone name: CITB-H1_2128F4

Summary Statistics
Consensus quality: 133990 bases at least Q40

Consensus quality: 137198 bases at least Q20
Consensus quality: 137811 bases at least Q20
Estimated insert size: 135000; pulse field gel estimation
Estimated insert size: 138488; sum-of-contigs estimation
Quality coverage: 6.51 in Q20 bases; pulse field gel estimation
Quality coverage: 6.35 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 99036: contig of 99036 bp in length
* 99037 99136: gap of unknown length
* 99137 107839: contig of 8703 bp in length
* 107840 107939: gap of unknown length
* 107940 110235: contig of 2296 bp in length
* 110236 110335: gap of unknown length
* 110336 138638: contig of 28303 bp in length.
Location/Qualifiers
1. 138638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2128F4"
/clone_lib="Caltech human BAC library D"
BASE COUNT 40604 a 25948 c 26675 g 45097 t 314 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AC010625/rev ..
Align seg 1/1 to reverse of: AC010625 from: 1 to: 138638
977 PhleuProAProAlaUValArg 985
|||||
51635 TTTCTACGATCTGCTCTGTGAGA 51609
seq_name: gb_htg:AC091917
seq_documentation_block:
LOCUS AC091917 145022 bp DNA HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-231G15, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
AC091917
AC091917.1 GI:14333853
VERSION HTG: HTGS-PHASE1: HTGS-DRAFT.
KEYWORDS human.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 145022)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 145022)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 495228
Center clone name: RPCI-11_231G15

Summary Statistics
Consensus quality: 131499 bases at least Q40
Consensus quality: 139142 bases at least Q20
Consensus quality: 140446 bases at least Q20
Estimated insert size: 162640; agarose-fp estimation
Quality coverage: 4.31 in Q20 bases; sum-of-contigs estimation
Quality coverage: 4.88 in Q20 bases; agarose-fp estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2779: contig of 2779 bp in length
* 2780 2879: gap of unknown length
* 2880 5833: contig of 2954 bp in length
* 5834 5933: gap of unknown length
* 5934 8788: contig of 2855 bp in length
* 8789 8888: gap of unknown length
* 8889 10614: contig of 1726 bp in length
* 10615 10714: gap of unknown length
* 10715 15852: contig of 5138 bp in length
* 15853 15952: gap of unknown length
* 15953 19797: contig of 3845 bp in length
* 19798 19897: gap of unknown length
* 19898 24431: contig of 4534 bp in length
* 24432 24531: gap of unknown length
* 24532 31863: contig of 7332 bp in length
* 31864 31963: gap of unknown length
* 31964 42763: contig of 10800 bp in length
* 42764 42864: gap of unknown length
* 42864 51923: contig of 9060 bp in length
* 51924 52023: gap of unknown length
* 52024 65520: contig of 13497 bp in length
* 65521 65620: gap of unknown length
* 65621 77977: contig of 12357 bp in length
* 77978 78077: gap of unknown length
* 78078 94277: contig of 16200 bp in length
* 94278 94377: gap of unknown length
* 94378 111201: contig of 16824 bp in length
* 111202 111301: gap of unknown length
* 111302 145022: contig of 33721 bp in length.
Location/Qualifiers
1. 145022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-231G15"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 42292 a 27711 c 27476 g 46116 t 1427 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AC091917/rev ..
Align seg 1/1 to reverse of: AC091917 from: 1 to: 145022
977 PhleuProAProAlaUValArg 985
|||||

97476 TTCTACGAGTCTCTGTGTAAGA 97450

seq_name: gb_htg:AC027067

seq_documentation_block:
 LOCUS AC027067 146403 bp DNA HTG 19-APR-2000
 DEFINITION Homo sapiens chromosome 1 clone RP11-185F19 map 1, WORKING DRAFT
 SEQUENCE 15 unordered pieces.
 ACCESSION AC027067
 VERSION AC027067.2 GI:7596882
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 146403)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 1, clone RP11-185F19
 Unpublished
 2 (bases 1 to 146403)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
 Campolino,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hages,B., Hearford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,R., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldim,J., Menes,L., Mihova,T., Miranda,C., Mienna V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Plerke,N.,
 Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testave,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 COMMENT
 Submitted (27-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 19, 2000 this sequence version replaced gi:7330330.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L7613
 Center clone name: 185_F_19
 Summary Statistics
 Sequencing vector: M13, M7815, 100% of reads
 Chemistry: Dye-terminator Big Dye, 100% of reads
 Assembly program: Phrap, version 0.960731
 Consensus quality: 138345 bases at least Q40
 Consensus quality: 142293 bases at least Q30
 Consensus quality: 143929 bases at least Q20
 Insert size: 139000; agarose-fp
 Insert size: 145003; sum-of-contigs
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 2016 2115: contig of 2015 bp in length
 2116 4862: contig of 2747 bp in length
 4863 4962: gap of 100 bp
 4963 7338: contig of 2376 bp in length
 7339 7438: gap of 100 bp
 7439 12966: contig of 5528 bp in length
 12967 13066: gap of 100 bp
 13067 19946: contig of 6880 bp in length
 19947 20046: gap of 100 bp
 20047 26200: contig of 6154 bp in length
 26201 26300: gap of 100 bp
 26301 33247: contig of 6947 bp in length
 33248 33347: gap of 100 bp
 33348 41169: contig of 7822 bp in length
 41170 41269: gap of 100 bp
 41270 52812: contig of 11543 bp in length
 52813 52912: gap of 100 bp
 52913 64180: contig of 11268 bp in length
 64181 64280: gap of 100 bp
 64281 76545: contig of 12265 bp in length
 76546 76645: gap of 100 bp
 76646 90751: contig of 14106 bp in length
 90752 90851: gap of 100 bp
 90852 105281: contig of 14430 bp in length
 105282 105381: gap of 100 bp
 105382 119852: contig of 14471 bp in length
 119853 119952: gap of 100 bp
 119953 146403: contig of 26451 bp in length.

FEATURES
 source
 1..146403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1"
 /clone="RP11-185F19"
 /clone_lib="RPCT-11 Human Male BAC"
 1..2015
 /note="assembly-fragment"
 2116..4862
 /note="assembly-fragment"
 4963..7338
 /note="assembly-fragment"
 7439..12966
 /note="assembly-fragment"
 13067..19946
 /note="assembly-fragment"
 20047..26200
 /note="assembly-fragment"
 26301..33247
 /note="assembly-fragment"
 33348..41169
 /note="assembly-fragment"
 41270..52812
 /note="assembly-fragment"
 52913..64180
 /note="assembly-fragment"
 clone_end:SP6
 vector_side:right
 64281..76545
 /note="assembly-fragment"
 76646..90751
 /note="assembly-fragment"
 90852..105281
 /note="assembly-fragment"
 105382..119852
 /note="assembly-fragment"
 119953..146403
 /note="assembly-fragment"
 BASE COUNT 48672 a 24607 c 24578 g 47143 t 1403 others


```
/note="assembly_fragment:00824
fragment_chain:3"
misc_feature
/note="assembly_fragment:00651
fragment_chain:3"
misc_feature
/note="assembly_fragment:00697
fragment_chain:3"
misc_feature
82640..85615
/note="assembly_fragment:00073
fragment_chain:4"
misc_feature
85716..87818
/note="assembly_fragment:00975
fragment_chain:4"
misc_feature
87919..90087
/note="assembly_fragment:00466
fragment_chain:5"
misc_feature
90188..101960
/note="assembly_fragment:01442
fragment_chain:5"
misc_feature
102061..112215
/note="assembly_fragment:01023
fragment_chain:6"
misc_feature
112316..114961
/note="assembly_fragment:01196
fragment_chain:6"
misc_feature
115062..121907
/note="assembly_fragment:00161.0"
misc_feature
122008..127169
/note="assembly_fragment:00400"
misc_feature
127270..133320
/note="assembly_fragment:00608"
misc_feature
132421..142336
/note="assembly_fragment:00974"
misc_feature
142337..145085
/note="assembly_fragment:01009"
misc_feature
145186..148456
/note="assembly_fragment:01118"
BASE COUNT 38044 a 33998 c 34836 g 39257 t 2321 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AL133550/rev ..
Align seg 1/1 to reverse of: AL133550 from: 1 to: 148456
700 ValAlaGlySerLeuSerLeuValLeu 708
|||||
6954 GTTCTGTGAAACCTTCTCTGTTTG 6928
seq_name: gb_htg:AC091907
seq_documentation_block:
LOCUS AC091907 150037 bp DNA HTG 01-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-195C11, WORKING DRAFT
SEQUENCE, 2 ordered pieces.
ACCESSION AC091907
VERSION AC091907.2 GI:14579736
KEYWORDS HTG; HTGS-PHASE2; HTGS-DRAFT; HTGS-ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150037)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
```

```
REFERENCE 2 (bases 1 to 150037)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 1, 2001 this sequence version replaced gi:14333843.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 481304
Center clone name: RPC1-11_195C11
-----
Summary Statistics
Consensus quality: 146786 bases at least Q40
Consensus quality: 149501 bases at least Q30
Consensus quality: 149814 bases at least Q20
Estimated insert size: 155670; agarose-fp estimation
Estimated insert size: 149937; sum-of-contigs estimation
Quality coverage: 7.6 in Q20 bases; agarose-fp estimation
Quality coverage: 7.89 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 7907: contig of 7907 bp in length
7908 8007: gap of unknown length
8008 150037: contig of 142030 bp in length.
Location/Qualifiers
1..150037
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-195C11"
/clone_lib="RPC1 human BAC library 11"
BASE COUNT 50775 a 26351 c 25832 g 46979 t 100 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AC091907/rev ..
Align seg 1/1 to reverse of: AC091907 from: 1 to: 150037
682 LysIlePheSerSerAlaThrSerLeu 690
|||||
53806 AAAATATTTCCTCAGCCACAAAGTTTG 53780
seq_name: gb_pr:AL359713
seq_documentation_block:
LOCUS AL359713 152966 bp DNA PRI 01-MAY-2001
DEFINITION Human DNA sequence from clone RP11-95P3 on chromosome 6, complete
sequence.
ACCESSION AL359713
VERSION AL359713.25 GI:13938809
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE 1 (bases 1 to 152966)
AUTHORS Kimberley, A.
TITLE Direct Submission
JOURNAL Submitted (01-May-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
COMMENT On May 3, 2001 this sequence version replaced gi:13446455.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
Rp11-95p3 is from the library RPCT-11.1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pACE3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-95p3. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone Rp11-95p3 is at 1 in this sequence. The
true left end of clone Rp1-73M23 is at 152867 in this sequence. The
true right end of clone Rp11-40E20 is at 17700 in this sequence.
Location/Qualifiers
1..152966
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Rp11-95p3"
/clone_lib="RPCT-11.1"
19..544
/note="match: GSS: Em:A2517849"
24..379
/note="match: GSS: Em:B63526"
32..494
/note="match: GSS: Em:AQ315706"
70..589
/note="match: GSS: Em:AQ285677"
179..362
/note="WIR repeat: matches 65..262 of consensus"
521..582
/note="L2 repeat: matches 2683..2741 of consensus"
720..1022
/note="Alusx repeat: matches 1..303 of consensus"
1327..1448
/note="AluDo/FRAM repeat: matches 181..298 of consensus"
2264..2575
/note="Alusg repeat: matches 1..311 of consensus"
3014..3093
/note="Alusg repeat: matches 1..311 of consensus"
3014..3296
/note="match: STS: Em:G27290"
5024..5477
/note="L1M1 repeat: matches 1012..1598 of consensus"
5581..5859
/note="Alusg repeat: matches 18..294 of consensus"
5986..6355
repeat_region

repeat_region
6367..6515
/note="L1MC repeat: matches 2228..2259 of consensus"
/note="FLAM.C repeat: matches 1..143 of consensus"
6971..7198
/note="L2 repeat: matches 1512..1735 of consensus"
8403..8838
/note="LTR7 repeat: matches 1..445 of consensus"
8799..9441
/note="match: GSS: Em:B54691"
complement(9393..9853)
/note="match: GSS: Em:AQ285817"
10311..10437
/note="L2 repeat: matches 2612..2747 of consensus"
10710..11022
/note="AluDo repeat: matches 1..309 of consensus"
12006..12156
/note="L2 repeat: matches 2594..2748 of consensus"
12182..13011
/note="L2 repeat: matches 1374..2344 of consensus"
13404..13631
/note="L1P3 repeat: matches 5054..5133 of consensus"
13479..13630
/note="L1P3 repeat: matches 5054..5133 of consensus"
13785..14820
/note="L1P3 repeat: matches 5054..5133 of consensus"
14928..16008
/note="L1P3 repeat: matches 5106..6149 of consensus"
16029..16110
/note="L1P3 repeat: matches 5054..5133 of consensus"
16478..16829
/note="L1MC3 repeat: matches 6189..6603 of consensus"
16868..17150
/note="L1MC3 repeat: matches 6189..6603 of consensus"
17246..17991
/note="Alusx repeat: matches 1..285 of consensus"
17992..18675
/note="L1MC4 repeat: matches 6608..7343 of consensus"
18676..18729
/note="MER67C repeat: matches 1..710 of consensus"
18731..19458
/note="L1MC4 repeat: matches 7343..7396 of consensus"
complement(18929..19542)
/note="MER4D repeat: matches 232..973 of consensus"
/note="match: GSS: Em:AQ342961"
complement(19031..19542)
/note="match: GSS: Em:AQ053552"
complement(19065..19495)
/note="match: GSS: Em:AQ545892"
complement(19130..19546)
/note="match: GSS: Em:B94529"
19265..19495
/note="match: GSS: Em:AQ373351 Em:AQ375803
match: STS: Em:G59266"
complement(19269..19546)
/note="match: GSS: Em:AQ663911"
complement(19325..19546)
/note="match: GSS: Em:AQ021494"
complement(19337..19501)
/note="match: GSS: Em:AQ268095"
complement(19370..19546)
/note="match: GSS: Em:AQ128461"
19566..19739
/note="Alusg/x repeat: matches 126..295 of consensus"
19740..20100
/note="L1M3 repeat: matches 7391..7739 of consensus"
21121..21429
/note="AluDo repeat: matches 1..310 of consensus"
22389..22728
/note="L2 repeat: matches 1..310 of consensus"
22591..22670
/note="L2 repeat: matches 1..310 of consensus"
22686..22761
/note="L2 repeat: matches 1..310 of consensus"
23098..23393
repeat_region


```

repeat_region      /note="AluIb repeat: matches 4. .298 of consensus"
23468. .23578
repeat_region      /note="L2 repeat: matches 2580. .2700 of consensus"
23980. .24292
misc_feature       /note="AluSg repeat: matches 2. .310 of consensus"
complement(27208. .27677)
repeat_region      /note="match: GSS: Em:AQ702871"
27819. .27957
repeat_region      /note="AluIb repeat: matches 163. .299 of consensus"
28701. .28930
repeat_region      /note="MLTIG repeat: matches 32. .301 of consensus"
29080. .29171
repeat_region      /note="MLTIG repeat: matches 433. .526 of consensus"
29303. .29711
repeat_region      /note="L2 repeat: matches 2323. .2710 of consensus"
29712. .29834
repeat_region      /note="MLTIG repeat: matches 29. .147 of consensus"
30140. .30547
repeat_region      /note="L2 repeat: matches 1806. .2252 of consensus"
complement(30720. .31207)
misc_feature       /note="match: STS: Em:HSJL17"
30919. .31120
repeat_region      /note="MERS8A repeat: matches 1. .208 of consensus"
32109. .32472
repeat_region      /note="RHEIC repeat: matches 1. .371 of consensus"
33360. .33511
repeat_region      /note="MIR repeat: matches 98. .250 of consensus"
34031. .34331
repeat_region      /note="AluSx repeat: matches 1. .297 of consensus"
34621. .34732
repeat_region      /note="MIR repeat: matches 120. .232 of consensus"
34888. .35007
repeat_region      /note="L2 repeat: matches 1836. .1959 of consensus"
35164. .35461
repeat_region      /note="AluSg repeat: matches 2. .298 of consensus"
35532. .35990
repeat_region      /note="L2 repeat: matches 2254. .2710 of consensus"
36631. .36790
repeat_region      /note="L2 repeat: matches 2342. .2501 of consensus"
37251. .37544
repeat_region      /note="AluSx repeat: matches 1. .294 of consensus"
38446. .38493
repeat_region      /note="L2 copies 4 mer caca 75% conserved"
38467. .38492

alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AL359713 ..

Align seg 1/1 to: AL359713 from: 1 to: 152966

451  AlAGIvARGATGleuSerleuLeu 459
|||||
54277 GCTGCCGCCGCTCAGCTCCTCT 54303

seq_name: gb_dr:AC004074

seq_documentation_block:
LOCUS      AC004074      153578 bp      DNA      PRI      12-APR-1998
DEFINITION Homo Sapiens Chromosome X clone bXMD759, complete sequence.
ACCESSION      AC004074
VERSION      AC004074.1 GI:3046270
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 153578)
AUTHORS      Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and

```

```

TITLE      Mazzarella,R.
JOURNAL      Direct Submission
REFERENCE      Unpublished (1997)
AUTHORS      2 (bases 1 to 153578)
BROWNSTEIN,B.H., States,D.J. and Mazzarella,R.
TITLE      Direct Submission
JOURNAL      Submitted (29-JAN-1998) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
REFERENCE      3 (bases 1 to 153578)
AUTHORS      Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE      Direct Submission
JOURNAL      Submitted (12-APR-1998) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
On Apr 13, 1998 this sequence version replaced gi:2822135.
COMMENT      Current status of this project is available at:
'http://www.lbc.wustl.edu/cgm/seq_projects.html'
Submitted by:
Elison Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 94404 USA
e-mail: ellison@genseq.apldbio.com

and

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu

and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@ibc.wustl.edu.

FEATURES
Source      1.153578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="bXMD759"
/chromosome="X"

BASE COUNT      45069 a 29281 c 29096 g 50132 t

ORIGIN

alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AC004074 ..

Align seg 1/1 to: AC004074 from: 1 to: 153578

848  VallysleuSerlleuAspleuSer 856
|||||
119704 GTTAAGTTGTCATTTAGATCTTCC 119730

seq_name: gb_htg:AL591908

seq_documentation_block:
LOCUS      AL591908      154715 bp      DNA      HTG      06-JUN-2001
DEFINITION Homo sapiens chromosome X clone RP11-246310, *** SEQUENCING IN
PROGRESS ***. 12 unordered pieces.

```

```

ACCESSION      AL591908
VERSION        AL591908.1
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL        1 (bases 1 to 154715)
REFERENCE      McIay,K.
TITLE          Direct Submission
AUTHORS        Submitted (03-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL        CB10 ISA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
                requests: clonerequests@sanger.ac.uk
COMMENT        -----
                Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk
                -----
                Project Information
                Center project name: BA246J10
                -----
                Summary Statistics
                Assembly Program: XGAP4; Version 4.5
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Consensus quality: 147425 bases at least Q40
                Consensus quality: 152715 bases at least Q30
                Insert size: 153615; sum-of-contigs
                Insert size: 170586; 2.4% error; agarose-fp
                Quality coverage: 3.97x in Q20 bases; sum-of-contigs quality
                coverage: 3.73x in Q20 bases; agarose-fp
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 12 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                -----
                1 10589: contig of 10589 bp in length
                * 10590 10689: gap of 100 bp
                * 23464: contig of 12775 bp in length
                * 23465 23564: gap of 100 bp
                * 23565 45151: contig of 21587 bp in length
                * 45152 45251: gap of 100 bp
                * 45252 50227: contig of 4976 bp in length
                * 50228 50327: gap of 100 bp
                * 50328 63924: contig of 13537 bp in length
                * 63925 64024: gap of 100 bp
                * 64025 84910: contig of 20886 bp in length
                * 84911 85010: gap of 100 bp
                * 85011 101389: contig of 16379 bp in length
                * 101390 101489: gap of 100 bp
                * 101489 113480: contig of 11991 bp in length
                * 113481 113580: gap of 100 bp
                * 113581 118940: contig of 5360 bp in length
                * 118941 119040: gap of 100 bp
                * 119041 139138: contig of 20098 bp in length
                * 139139 139238: gap of 100 bp
                * 139239 147328: contig of 8090 bp in length
                * 147329 147428: gap of 100 bp
                * 147429 154715: contig of 7287 bp in length.
                Location/Qualifiers
                1..154715
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="X"
                /clone_id="RP11-246J10"
                /clone_1lb="RPCT-11.1"
                1..10589
                /note="assembly fragment:01304
                clone_end:SP6

```

```

misc_feature      misc_side:left"
                    10690..23466
                    /note="assembly-fragment:00666
                    fragment_chain:1"
                    23565..45151
                    /note="assembly-fragment:00281
                    fragment_chain:1"
                    45252..50227
                    /note="assembly-fragment:01309
                    fragment_chain:1"
                    50328..63924
                    /note="assembly-fragment:00188"
                    64025..84910
                    /note="assembly-fragment:00634"
                    85011..101389
                    /note="assembly-fragment:00668"
                    101490..113480
                    /note="assembly-fragment:01065"
                    113581..118940
                    /note="assembly-fragment:01209"
                    119041..139138
                    /note="assembly-fragment:01266"
                    139239..147328
                    /note="assembly-fragment:01246"
                    147429..154715
                    /note="assembly-fragment:01430"
BASE COUNT      44837 a 31500 c 31007 g 46270 t 1101 others
ORIGIN
alignment_scores:
    Quality:      9.00          Length:      9
    Ratio:        1.000         Gaps:       0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AL591908/rev ..

Align seg 1/1 to reverse of: AL591908 from: 1 to: 154715

700 VALAAGlyserleuSerLeuValIlu 708
|||||TTTTTTTTTTTTTTTTTTTTTTT
4431 GTTGCTGGAGACCTTCTCTGTTTG 4405

seq_name: gb:htg:AC023214

seq_documentation_block:
LOCUS      AC023214   155691 bp      DNA
DEFINITION Homo sapiens clone RP11-313B12, WORKING DRAFT SEQUENCE, 12
            unordered pieces.
ACCESSION  AC023214
VERSION    AC023214.3   GI:8076865
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 155691)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE      Human Genome Project International Consortium
JOURNAL
REFERENCE  2 (bases 1 to 155691)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavsky,I., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
            Choedel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
            Dearlano,K., Dewar,K., Domingo,M., Doyle,M., Feneslor,J.,
            Ferreira,P., FitzHugh,W., Forrest.C., Gage,D., Galagan,J.,
            Garyyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lands,T., Lehoczeky,J., Levine,C., Liu,G., Locke,K.,
            Macdonald,P., Margulis,N., McEwan,P., McGuirk,A., McKernan,K.,
            McPheters,R., Meldrim,J., Menus.L., Morrow,J., Naylor.J.,

```

Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
 Pierre,N., Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
 Stojanovic,N., Sudramanian,A., Talama,J., Testaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.

JOURNAL

COMMENT

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:7139013.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L6599

Center clone name: 313_B_12

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 149532 bases at least Q40

Consensus quality: 152862 bases at least Q40

Insert size: 150000; agarose-gel

Quality coverage: 5.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 12 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1001: contig of 1001 bp in length
 1002 1101: gap of 100 bp
 1102 2737: contig of 1636 bp in length
 2738 2837: gap of 100 bp
 2838 4417: contig of 1580 bp in length
 4418 4517: gap of 100 bp
 4518 7080: contig of 2563 bp in length
 7081 7180: gap of 100 bp
 7181 9985: contig of 2805 bp in length
 9986 10085: gap of 100 bp
 10086 14330: contig of 4245 bp in length
 14331 14430: gap of 100 bp
 14431 16528: contig of 2098 bp in length
 16529 16628: gap of 100 bp
 16629 21294: contig of 4666 bp in length
 21295 21394: gap of 100 bp
 21395 26849: contig of 5455 bp in length
 26850 26949: gap of 100 bp
 26950 34876: contig of 7927 bp in length
 34877 34976: gap of 100 bp
 34977 81758: contig of 46782 bp in length
 81759 81858: gap of 100 bp
 81859 155691: contig of 73833 bp in length.

FEATURES

source

1. 155691
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

misc_feature

1. 1001
 /clone_lib="RP11 Human Male BAC"
 /note="assembly_fragment"

misc_feature

1102. 2737
 /note="assembly_fragment"

misc_feature /note="assembly_fragment"
 2838..4417
 /note="assembly_fragment"
 misc_feature 4518..7080
 /note="assembly_fragment"
 misc_feature 7181..9985
 /note="assembly_fragment"
 misc_feature 10086..14330
 /note="assembly_fragment"
 misc_feature 14431..16528
 /note="assembly_fragment"
 misc_feature 16629..21294
 /note="assembly_fragment"
 misc_feature 21395..26849
 /note="assembly_fragment"
 misc_feature 26950..34876
 /note="assembly_fragment"
 misc_feature 34977..81758
 /note="assembly_fragment"
 misc_feature 81859..155691
 /note="assembly_fragment"
 BASE COUNT 43630 a 31999 c 32619 g 46340 t 1103 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AC023214 ..

Align seg 1/1 to: AC023214 from: 1 to: 155691

892 ValGlnGlySerLeuSerLeuLeu 900

80755 GTTCAGGTTCTCATCTCTCTGCTG 80781

seq_name: gb_htg:AC022834

seq_documentation_block:

LOCUS AC022834 155833 bp DNA HTG 04-OCT-2000
 DEFINITION Homo sapiens chromosome 14 clone RP11-152N21 map 14, WORKING DRAFT
 SEQUENCE 16 unordered pieces.

ACCESSION AC022834
 VERSION AC022834.3 GI:10567661

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 155833)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 14, clone RP11-152N21
 Unpublished
 2 (bases 1 to 155833)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Bouckgaltier,B., Brown,A., Burkett,G., Castle,A.,
 Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Garayna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczyk,J., Levine,R., Liu,C., Liu,G., Locke,K.,
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrim,J., Meheus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,

JOURNAL COMMENT

Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 4, 2000 this sequence version replaced gi:7717129.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: submissions@genome.wi.mit.edu

Project Information

Center project name: 152_N_21
Center clone name: 16142
Sequencing Vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 148061 bases at least Q40
Consensus quality: 152026 bases at least Q30
Consensus quality: 153405 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 154333; sum-of-ctrls
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      2233: contig of 2233 bp in length
*      2234 2333: gap of 100 bp
*      2334 3352: contig of 1019 bp in length
*      3353 3452: gap of 100 bp
*      3453 6181: contig of 2729 bp in length
*      6182 6281: gap of 100 bp
*      6282 12746: contig of 6465 bp in length
*      12747 12846: gap of 100 bp
*      12847 18740: contig of 5894 bp in length
*      18741 18840: gap of 100 bp
*      18841 27275: contig of 8435 bp in length
*      27276 27375: gap of 100 bp
*      27376 36356: contig of 8981 bp in length
*      36357 36456: gap of 100 bp
*      36457 45635: contig of 9179 bp in length
*      45636 45735: gap of 100 bp
*      45736 36054: contig of 10319 bp in length
*      56055 56154: gap of 100 bp
*      56155 65246: contig of 9092 bp in length
*      65247 65346: gap of 100 bp
*      65347 75394: contig of 10048 bp in length
*      75395 75494: gap of 100 bp
*      75495 87010: contig of 11516 bp in length
*      87011 87110: gap of 100 bp
*      87111 111821: contig of 24711 bp in length
*      111822 111921: gap of 100 bp
*      111922 127551: contig of 15630 bp in length
*      127552 127651: gap of 100 bp
*      127652 143258: contig of 15607 bp in length
*      143259 143358: gap of 100 bp
*      143359 155833: contig of 12475 bp in length.

```

FEATURES source

Location/Qualifiers
1..155833

```

misc-feature /organism="Homo sapiens"
misc-feature /db_xref="taxon:9606"
misc-feature /chromosome="14"
misc-feature /map="14"
misc-feature /clone="RP11-152N21"
misc-feature /clone_lib="RPCT-11 Human Male BAC"
misc-feature 1..2233
misc-feature /note="assembly_fragment"
misc-feature clone_end:SP6
misc-feature vector_side:left"
misc-feature 2334..3352
misc-feature /note="assembly_fragment"
misc-feature 3453..6181
misc-feature /note="assembly_fragment"
misc-feature 6282..12746
misc-feature /note="assembly_fragment"
misc-feature 12847..18740
misc-feature /note="assembly_fragment"
misc-feature 18841..27275
misc-feature /note="assembly_fragment"
misc-feature 27376..36356
misc-feature /note="assembly_fragment"
misc-feature 36457..45635
misc-feature /note="assembly_fragment"
misc-feature 45736..56054
misc-feature /note="assembly_fragment"
misc-feature 56155..65246
misc-feature /note="assembly_fragment"
misc-feature 65347..75394
misc-feature /note="assembly_fragment"
misc-feature 75495..87010
misc-feature /note="assembly_fragment"
misc-feature 87111..111821
misc-feature /note="assembly_fragment"
misc-feature 111922..127551
misc-feature /note="assembly_fragment"
misc-feature 127652..143258
misc-feature /note="assembly_fragment"
misc-feature 143359..155833
misc-feature /note="assembly_fragment"
misc-feature clone_end:R7
misc-feature vector_side:right"

```

BASE COUNT 45833 a 32120 c 32332 g 43969 t 1579 others
ORIGIN

alignment_scores: Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AC022834 ..
Align seg 1/1 to: AC022834 from: 1 to: 155833

892 ValGInGlySerLeuSerLeuLeu 900
|||||
44565 GTTCAAGTTCACTATCTCGTGTGCTG 44591

seq_name: gb_hhg:AP001444

seq_documentation_block:

LOCUS AP001444 157978 bp DNA HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-638P2 map 11q13, WORKING
DRAFT SEQUENCE, 45 unordered pieces.
ACCESSION AP001444
VERSION AP001444.2 GI:8117318
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-638P2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 157978)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 157,978 genomic DNA of 11q13
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 157978)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Katsushika Univ., 1-15-1 Katsushika, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gscc.riken.go.jp,
URL:http://hqp.gscc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7262577.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hqp.gscc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: HumDrat11
Center clone name: RP11-638P2
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 13313 bases at least Q40
Consensus quality: 14506 bases at least Q30
Consensus quality: 15069 bases at least Q20
Insert size: 153578; sum-of-contigs
Quality coverage: 4.15x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
45 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.
1
12893 12792 contig of 12792 bp in length
20206 20105 contig of 7213 bp in length
29354 29253 contig of 9048 bp in length
36422 36321 contig of 6968 bp in length
42046 41945 contig of 5524 bp in length
49996 49895 contig of 7850 bp in length
54496 54395 contig of 4400 bp in length
58404 58303 contig of 3808 bp in length
62036 62035 contig of 3632 bp in length
66136 66135 contig of 4380 bp in length
70939 70838 contig of 4223 bp in length
75805 75704 contig of 4766 bp in length
80466 80365 contig of 4561 bp in length
84452 84351 contig of 3886 bp in length
88266 88265 contig of 3814 bp in length
92478 92477 contig of 4112 bp in length
96940 96939 contig of 4255 bp in length
101395 101394 contig of 4255 bp in length
104895 104794 contig of 3400 bp in length
107972 107871 contig of 2977 bp in length
111199 111098 contig of 3127 bp in length
116712 116711 contig of 2195 bp in length
119910 119909 contig of 3218 bp in length
120010 120009 contig of 3098 bp in length
122357 122356 contig of 2247 bp in length
125506 125505 contig of 3150 bp in length
133156 133155 contig of 2877 bp in length
136376 136375 contig of 2053 bp in length
138104 138103 contig of 1403 bp in length
1628 bp in length

138205 139803 contig of 1599 bp in length
139904 141655 contig of 1752 bp in length
141756 143702 contig of 1947 bp in length
143803 145350 contig of 1548 bp in length
145451 147233 contig of 1783 bp in length
147334 149496 contig of 2163 bp in length
149597 150204 contig of 608 bp in length
150305 151726 contig of 1422 bp in length
151827 152990 contig of 1164 bp in length
153091 154455 contig of 1365 bp in length
154556 155671 contig of 1186 bp in length
155772 156791 contig of 1020 bp in length
156892 157978 contig of 1087 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
12793 12892: contig of 12792 bp in length
12793 12892: gap of 100 bp
12893 20105: contig of 7213 bp in length
20106 20205: gap of 100 bp
20206 29253: contig of 9048 bp in length
29254 29353: gap of 100 bp
29354 36321: contig of 6968 bp in length
36322 36421: gap of 100 bp
36422 41945: contig of 5524 bp in length
41946 42045: gap of 100 bp
42046 49895: contig of 7850 bp in length
49896 49995: gap of 100 bp
49996 54395: contig of 4400 bp in length
54396 54495: gap of 100 bp
54496 58303: contig of 3808 bp in length
58304 58403: gap of 100 bp
58404 62035: contig of 3632 bp in length
62036 62135: gap of 100 bp
62136 66515: contig of 4380 bp in length
66516 66615: gap of 100 bp
66616 70838: contig of 4223 bp in length
70839 70938: gap of 100 bp
70939 75704: contig of 4766 bp in length
75705 75804: gap of 100 bp
75805 80365: contig of 4561 bp in length
80366 80465: gap of 100 bp
80466 84351: contig of 3886 bp in length
84352 84451: gap of 100 bp
84452 88265: contig of 3814 bp in length
88266 88365: gap of 100 bp
88366 92477: contig of 4112 bp in length
92478 92577: gap of 100 bp
92578 96939: contig of 4362 bp in length
96940 97039: gap of 100 bp
97040 101294: contig of 4255 bp in length
101295 101394: gap of 100 bp
101395 104794: contig of 3400 bp in length
104795 104894: gap of 100 bp
104895 107871: contig of 2977 bp in length
107872 107971: gap of 100 bp
107972 111098: contig of 3127 bp in length
111099 111198: gap of 100 bp
111199 113393: contig of 2195 bp in length
113394 113493: gap of 100 bp
113494 116711: contig of 3218 bp in length
116712 116811: gap of 100 bp
116812 119909: contig of 3098 bp in length
119910 120009: gap of 100 bp
120010 122256: contig of 2247 bp in length
122257 122356: gap of 100 bp
122357 125506: contig of 3150 bp in length

```

* 125507 125606: gap of 100 bp
* 125607 127925: contig of 2319 bp in length
* 127926 128025: gap of 100 bp
* 128026 130902: contig of 2877 bp in length
* 130903 131002: gap of 100 bp
* 131003 133055: contig of 2053 bp in length
* 133056 133155: gap of 100 bp
* 133156 134558: contig of 1403 bp in length
* 134559 134658: gap of 100 bp
* 134659 136376: contig of 1718 bp in length
* 136377 136476: gap of 100 bp
* 136477 138104: contig of 1628 bp in length
* 138105 138204: gap of 100 bp
* 138205 139603: contig of 1599 bp in length
* 139604 139903: gap of 100 bp
* 139904 141655: contig of 1752 bp in length
* 141656 141755: gap of 100 bp
* 141756 143702: contig of 1947 bp in length
* 143703 143802: gap of 100 bp
* 143803 145350: contig of 1548 bp in length
* 145351 145450: gap of 100 bp
* 145451 147233: contig of 1783 bp in length
* 147234 147333: gap of 100 bp
* 147334 149496: contig of 2163 bp in length
* 149497 149596: gap of 100 bp
* 149597 150204: contig of 608 bp in length
* 150205 150304: gap of 100 bp
* 150305 151726: contig of 1422 bp in length
* 151727 151826: gap of 100 bp
* 151827 152890: contig of 1164 bp in length
* 152891 153090: gap of 100 bp
* 153091 154453: contig of 1365 bp in length
* 154454 154553: gap of 100 bp
* 154554 155671: contig of 1116 bp in length
* 155672 155771: gap of 100 bp
* 155772 156791: contig of 1020 bp in length
* 156792 156891: gap of 100 bp
* 156892 157978: contig of 1087 bp in length.

```

FEATURES
source
1. 157978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q13"

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AP001444/rev ..

Align seg 1/1 to reverse of: AP001444 from: 1 to: 157978

827 LeuValSerCysCysLeuSerAlaAsn 835

117486 CTAGTGTCTGCTGCTGCTGCTAAT 117460

seq_name: gb_hlcg:AC055718

seq_documentation_block:

LOCUS AC055718 158746 bp DNA HTG 11-NOV-2000

DEFINITION Homo sapiens chromosome 3 clone RP11-554B20, WORKING DRAFT

ACCESSION AC055718

VERSION AC055718.15 GI:11136656

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

ORGANISM human.

REFERENCE 1 (bases 1 to 158746)

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Isman, F.R., Allen, C., Alshrocks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonini, D., Bouck, J., Bowles, S., Brivaga, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H., Douhaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, K., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, T., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteberg, O., Lieu, C., Liu, J., Liu, W., Louisedge, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nelson, E., Nwokenkwo, S., Ogutu, M., Okwono, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Pimus, E., Pu, L.L., Qulles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Roile, M., Ruiz, S., Saverly, G., Scherer, S., Scott, E., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, A., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Washington, S., Williams, G., Williamson, A., Wleczky, R., Woodan, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

TITLE

Direct Submission
Unpublished
2 (bases 1 to 158746)
Worley, K.C.
Direct Submission
Submitted (18-Apr-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 10, 2000 this sequence version replaced gi:10121940.

COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HANUM

Center clone name: RP11-554B20

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 99% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 140809 bases at least Q40

Consensus quality: 150659 bases at least Q30

Consensus quality: 155250 bases at least Q20

Estimated insert size: 154907; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
44955: contig of 44955 bp in length
45056: gap of unknown length
45056: contig of 35495 bp in length
80551: gap of unknown length
80551: contig of 20339 bp in length
100990: gap of unknown length
101090: contig of 13794 bp in length
114883: gap of unknown length
114884: contig of 14701 bp in length
129684: gap of unknown length
129685: contig of 11358 bp in length
141143: gap of unknown length
141243: contig of 7896 bp in length
149139: gap of unknown length
149239: contig of 4490 bp in length
153729: gap of unknown length
153829: contig of 2040 bp in length
153869: gap of unknown length
153968: contig of 1444 bp in length
157413: gap of unknown length
157513: contig of 1234 bp in length.

FEATURES
Source      Location/Qualifiers
1..158746   Homo sapiens
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-554B20"

BASE COUNT  48676 a 30025 c 30078 g 48948 t 1019 others

ORIGIN
420 ValLeuLeuThrThrGlyLeuLeuCys 428
|||||
793 GTGCTCCTAACCACTGACACTTTGT 819

alignment_scores:
Quality: 9.00      Length: 9
Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AC055718 ..

Align seg 1/1 to: AC055718 from: 1 to: 158746

420 ValLeuLeuThrThrGlyLeuLeuCys 428
|||||
793 GTGCTCCTAACCACTGACACTTTGT 819

seq_name: gb_htg:AC011838

seq_documentation_block:
LOCUS      AC011838      158807 bp      DNA      HTG      09-SEP-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-21102 map 8, *** SEQUENCING IN
PROGRESS ***, 28 unordered pieces.
ACCESSION  AC011838
VERSION    AC011838.3 GI:10045320
KEYWORDS   HTG: HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 158807)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            Homo sapiens chromosome 8, clone RP11-21102
            Unpublished
            2 (bases 1 to 158807)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boucknight,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donejan,L., Doyle,M.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,

```

```

TITLE
JOURNAL
COMMENT
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
Lenczky,J., Lien,C., Locke,K., MacDonald,P., Margolis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliou,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced g1:6479089.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RV/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2598
Center clone name: 211_O_2
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1360 1459: gap of 100 bp
1360 1459: contig of 1359 bp in length
1460 2767: gap of 100 bp
1460 2767: contig of 1307 bp in length
2867 5284: gap of 100 bp
2867 5284: contig of 2418 bp in length
5285 5384: gap of 100 bp
5285 5384: contig of 2599 bp in length
5385 7983: gap of 100 bp
5385 7983: contig of 2599 bp in length
7984 8083: gap of 100 bp
7984 8083: contig of 2061 bp in length
8084 10144: gap of 100 bp
8084 10144: contig of 2061 bp in length
10145 10244: gap of 100 bp
10145 10244: contig of 1385 bp in length
10245 11629: gap of 100 bp
10245 11629: contig of 1385 bp in length
11630 11729: gap of 100 bp
11630 11729: contig of 2299 bp in length
11730 14028: gap of 100 bp
11730 14028: contig of 2299 bp in length
14029 14128: gap of 100 bp
14029 14128: contig of 2997 bp in length
14129 17125: gap of 100 bp
14129 17125: contig of 2997 bp in length
17126 17225: gap of 100 bp
17126 17225: contig of 2912 bp in length
17226 20137: gap of 100 bp
17226 20137: contig of 2912 bp in length
20138 20237: gap of 100 bp
20138 20237: contig of 3566 bp in length
20238 23803: gap of 100 bp
20238 23803: contig of 3566 bp in length
23804 23903: gap of 100 bp
23804 23903: contig of 4018 bp in length
23904 27921: gap of 100 bp
23904 27921: contig of 4018 bp in length
27922 28021: gap of 100 bp
27922 28021: contig of 5009 bp in length
28022 33030: gap of 100 bp
28022 33030: contig of 5009 bp in length
33031 33130: gap of 100 bp
33031 33130: contig of 4283 bp in length
33131 37413: gap of 100 bp
33131 37413: contig of 4283 bp in length
37414 37513: gap of 100 bp
37414 37513: contig of 3364 bp in length
40878 40977: gap of 100 bp
40878 40977: contig of 5524 bp in length
40978 46301: gap of 100 bp
40978 46301: contig of 5524 bp in length
46302 46401: gap of 100 bp
46302 46401: contig of 5179 bp in length
46402 51580: gap of 100 bp
46402 51580: contig of 5179 bp in length
51581 51680: gap of 100 bp
51581 51680: contig of 4325 bp in length
56006 56105: gap of 100 bp
56006 56105: contig of 5214 bp in length
56106 61319: gap of 100 bp
56106 61319: contig of 5214 bp in length
61320 61419: gap of 100 bp
61320 61419: contig of 6431 bp in length
61420 67850: gap of 100 bp
61420 67850: contig of 6431 bp in length
67851 74230: gap of 100 bp
67851 74230: contig of 6280 bp in length
74231 74330: gap of 100 bp
74231 74330: contig of 6280 bp in length

```

	*	74331	81281:	contig of 6591 bp	in length
	*	81282	81381:	gap of 100 bp	
	*	81382	88330:	contig of 6549 bp	in length
	*	88331	88430:	gap of 100 bp	
	*	97676	97775:	contig of 9245 bp	in length
	*	97776	106600:	contig of 100 bp	
	*	106601	106700:	contig of 8825 bp	in length
	*	106701	106700:	gap of 100 bp	
	*	118048	118147:	contig of 11347 bp	in length
	*	118148	118016:	gap of 100 bp	
	*	130817	130916:	contig of 12669 bp	in length
	*	130917	144282:	gap of 100 bp	
	*	144283	144382:	contig of 13366 bp	in length
	*	144383	144382:	gap of 100 bp	
	*	144383	158807:	contig of 14425 bp	in length
FEATURES					
source					
			1..158807		
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/chromosome="8"		
			/map="8"		
			/clone="RP11-21102"		
			/clone_lib="RPCT-11 Human Male BAC"		
			1..1359		
misc_feature			/note="assembly_fragment"		
misc_feature			1460..2766		
misc_feature			/note="assembly_fragment"		
misc_feature			2861..5284		
misc_feature			/note="assembly_fragment"		
misc_feature			5385..7983		
misc_feature			/note="assembly_fragment"		
misc_feature			8084..10144		
misc_feature			/note="assembly_fragment"		
misc_feature			10245..11629		
misc_feature			/note="assembly_fragment"		
misc_feature			11730..14028		
misc_feature			/note="assembly_fragment"		
misc_feature			14129..17125		
misc_feature			/note="assembly_fragment"		
misc_feature			17226..20137		
misc_feature			/note="assembly_fragment"		
misc_feature			20238..23803		
misc_feature			/note="assembly_fragment"		
misc_feature			23904..27921		
misc_feature			/note="assembly_fragment"		
misc_feature			28022..33030		
misc_feature			/note="assembly_fragment"		
misc_feature			33131..37413		
misc_feature			/note="assembly_fragment"		
misc_feature			37514..40877		
misc_feature			/note="assembly_fragment"		
misc_feature			40978..46301		
misc_feature			/note="assembly_fragment"		
misc_feature			46402..51580		
misc_feature			/note="assembly_fragment"		
misc_feature			51681..56005		
misc_feature			/note="assembly_fragment"		
misc_feature			clone.end:T7		
misc_feature			vector_side:left		
misc_feature			56106..61319		
misc_feature			/note="assembly_fragment"		
misc_feature			clone.end:SP6		
misc_feature			vector_side:right		
misc_feature			61420..67850		
misc_feature			/note="assembly_fragment"		
misc_feature			67951..74230		
misc_feature			/note="assembly_fragment"		
misc_feature			74331..81281		
misc_feature			/note="assembly_fragment"		
misc_feature			81382..88330		
misc_feature			/note="assembly_fragment"		
misc_feature			88431..97675		
misc_feature			/note="assembly_fragment"		

```

misc_feature      97776..106600      /note="assembly_fragment"
misc_feature      106701..118047      /note="assembly_fragment"
misc_feature      118148..130816      /note="assembly_fragment"
misc_feature      130917..144282      /note="assembly_fragment"
misc_feature      144383..158807      /note="assembly_fragment"
BASE COUNT      48164 a 30675 c 30029 g 47222 t 2717 others
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AC011838/rev ...
Align seg 1/1 to reverse of: AC011838 from: 1 to: 158807
483 IieHrSerhrrTySerSerleu 491
|||||
133193 ATTACAAGCAGCTATTCACGCTTCTT 133167
seq_name: gb_htg:AC092361
seq_documentation_block:
LOCUS      AC092361      159082 bp      DNA      HTG      03-JUL-2001
DEFINITION      Homo sapiens chromosome 5 clone RP11-356021, WORKING DRAFT
SEQUENCE      30 unordered pieces.
AC092361
AC092361.1      GI:14589550
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      DOE Joint Genome Institute.
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 5
JOURNAL      Unpublished
2 (bases 1 to 159082)
DIRECT SUBMISSION
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 543426
Center clone name: RPC1-11_356021
-----
Summary Statistics
Consensus quality: 129435 bases at least Q40
Consensus quality: 148098 bases at least Q30
Consensus quality: 150725 bases at least Q20
Estimated insert size: 175410; agarose-1p estimation
Estimated insert size: 156182; sum-of-contigs estimation
Quality coverage: 5.67 in Q20 bases; agarose-1p estimation
Quality coverage: 6.37 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```


Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-09-697-089-2 x AC004993 ..

Align seg 1/1 to: AC004993 from: 1 to: 161453

517 CysleuLeuGlyLeuSerIleAlaIys 525
|||||
160371 TGTCTGCTAGGCTTGAGCATAGCCAA 160397

seq_name: gb_htg:AC018601

seq documentation block:

LOCUS AC018601 162030 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-28G15, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC018601
VERSION AC018601.3 GI:9137029
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 162030)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-28G15

Unpublished

2 (bases 1 to 162030)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boguslavsky,L., Bouhagalter,B., Brown,A., Burkett,G., Castle,A.,

Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,

Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galegan,J.,

Gardina,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehocsky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

Meltrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,

O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,

Sanos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,

Vassiliev,H., Viel,R., Vo,A., Wu,X., Wymann,D., Ye,W.J., Zimmer,A.

and Zody,M.

Direct Submission

Submitted (14-DEC-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6649328.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L4754

Center clone name: 28_G_15

* NOTE: This record contains 165 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 841: contig of 841 bp in length
842 941: gap of 100 bp
942 1821: contig of 880 bp in length
1822 1921: gap of 100 bp
1922 2799: contig of 878 bp in length
2800 2899: gap of 100 bp
2900 3805: contig of 906 bp in length
3806 3905: gap of 100 bp
3906 4798: contig of 893 bp in length
4799 4898: gap of 100 bp
4899 5807: contig of 909 bp in length
5808 5907: gap of 100 bp
5908 6834: contig of 927 bp in length
6835 6934: gap of 100 bp
6935 7842: contig of 908 bp in length
7843 7942: gap of 100 bp
7943 8824: contig of 882 bp in length
8825 8924: gap of 100 bp
8925 9783: contig of 859 bp in length
9784 9883: gap of 100 bp
9884 10760: contig of 877 bp in length
10761 10860: gap of 100 bp
10861 11719: contig of 859 bp in length
11720 11819: gap of 100 bp
11820 12722: contig of 903 bp in length
12723 12822: gap of 100 bp
12823 13727: contig of 905 bp in length
13728 13827: gap of 100 bp
13828 14717: contig of 890 bp in length
14718 14817: gap of 100 bp
14818 15713: contig of 896 bp in length
15714 15813: gap of 100 bp
15814 16700: contig of 887 bp in length
16701 16800: gap of 100 bp
16801 17620: contig of 820 bp in length
17621 17720: gap of 100 bp
17721 18588: contig of 868 bp in length
18589 18688: gap of 100 bp
18689 19570: contig of 882 bp in length
19571 19670: gap of 100 bp
19671 20513: contig of 843 bp in length
20514 20613: gap of 100 bp
20614 21482: contig of 869 bp in length
21483 21582: gap of 100 bp
21583 22454: contig of 872 bp in length
22455 22554: gap of 100 bp
22555 23433: contig of 879 bp in length
23434 23533: gap of 100 bp
23534 24419: contig of 886 bp in length
24420 24519: gap of 100 bp
24520 25419: contig of 900 bp in length
25420 25519: gap of 100 bp
25520 26392: contig of 873 bp in length
26393 26492: gap of 100 bp
26493 27381: contig of 889 bp in length
27382 27481: gap of 100 bp
27482 28315: contig of 834 bp in length
28316 28415: gap of 100 bp
28416 29298: contig of 883 bp in length
29299 29398: gap of 100 bp
29399 30300: contig of 902 bp in length
30401 30400: gap of 100 bp
30401 31267: contig of 867 bp in length
31268 31367: gap of 100 bp
31368 32252: contig of 885 bp in length
32253 32352: gap of 100 bp
32353 33217: contig of 865 bp in length
33218 33317: gap of 100 bp
33318 34192: contig of 875 bp in length
34193 34292: gap of 100 bp
34293 35187: contig of 895 bp in length
35188 35287: gap of 100 bp
35288 36190: contig of 903 bp in length

* 36191 36290: gap of 100 bp
 * 36291 37157: contig of 867 bp in length
 * 37158 37257: gap of 100 bp
 * 37258 38132: contig of 875 bp in length
 * 38133 38233: gap of 100 bp
 * 38233 39150: contig of 918 bp in length
 * 39151 39250: gap of 100 bp
 * 39251 40130: contig of 880 bp in length
 * 40131 40230: gap of 100 bp
 * 40231 41108: contig of 878 bp in length
 * 41109 41208: gap of 100 bp
 * 41209 42088: contig of 872 bp in length
 * 42081 42180: gap of 100 bp
 * 42181 43048: contig of 868 bp in length
 * 43049 43148: gap of 100 bp
 * 43149 44039: contig of 891 bp in length
 * 44040 44139: gap of 100 bp
 * 44140 45025: contig of 886 bp in length
 * 45026 45125: gap of 100 bp
 * 45126 46002: contig of 877 bp in length
 * 46003 46102: gap of 100 bp
 * 46103 47000: contig of 898 bp in length
 * 47001 47100: gap of 100 bp
 * 47101 47966: contig of 866 bp in length
 * 47967 48066: gap of 100 bp
 * 48067 48956: contig of 890 bp in length
 * 48957 49056: gap of 100 bp
 * 49057 49974: contig of 918 bp in length
 * 49975 50074: gap of 100 bp
 * 50075 50966: contig of 892 bp in length
 * 50967 51066: gap of 100 bp
 * 51067 51975: contig of 909 bp in length
 * 51976 52075: gap of 100 bp
 * 52076 52960: contig of 885 bp in length
 * 52961 53060: gap of 100 bp
 * 53061 53936: contig of 876 bp in length
 * 53937 54036: gap of 100 bp
 * 54037 54901: contig of 865 bp in length
 * 54902 55001: gap of 100 bp
 * 55002 55881: contig of 880 bp in length
 * 55882 55981: gap of 100 bp
 * 55982 56855: contig of 874 bp in length
 * 56856 56955: gap of 100 bp
 * 56956 57851: contig of 886 bp in length
 * 57852 57951: gap of 100 bp
 * 57952 58816: contig of 865 bp in length
 * 58817 58916: gap of 100 bp
 * 58917 59789: contig of 873 bp in length
 * 59790 59889: gap of 100 bp
 * 59890 60770: contig of 881 bp in length
 * 60771 60870: gap of 100 bp
 * 60871 61755: contig of 885 bp in length
 * 61756 61855: gap of 100 bp
 * 61856 62739: contig of 884 bp in length
 * 62740 62839: gap of 100 bp
 * 62840 63731: contig of 892 bp in length
 * 63732 63831: gap of 100 bp
 * 63832 64691: contig of 860 bp in length
 * 64692 64791: gap of 100 bp
 * 64792 65676: contig of 885 bp in length
 * 65677 65776: gap of 100 bp
 * 65777 66676: contig of 900 bp in length
 * 66677 66776: gap of 100 bp
 * 66777 67657: contig of 881 bp in length
 * 67658 67757: gap of 100 bp
 * 67758 68614: contig of 857 bp in length
 * 68615 68714: gap of 100 bp
 * 68715 69593: contig of 879 bp in length
 * 69594 69693: gap of 100 bp

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 us-09-697-089-2 x AC018601/rev ...
 Align seg 1/1 to reverse of: AC018601 from: 1 to: 162030
 682 LysillepheserSerA1arhserleu 690
 115408 AAAATATTTTCACGCCACAGTTTG 115382
 seq_name: gb_hgt:AP002986
 seq_documentation_block:
 LOCUS AP002986 162066 bp DNA HTG 05-DEC-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-121M22 map 11q, WORKING DRAFT
 SEQUENCE, 41 unordered pieces.
 ACCESSION AP002986
 VERSION AP002986.1 GI:11559301
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens DNA, clone:RP11-121M22.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 162066)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Homo sapiens 162,066 genomic DNA of 11q
 JOURNAL Published Only in Database (2000) In press
 REFERENCE 2 (bases 1 to 162066)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 COMMENT
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDrafc11
 Center clone name: RP11-121M22
 ----- Summary Statistics
 Sequencing vector: PCR products: 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 150730 bases at least Q40
 Consensus quality: 155539 bases at least Q30
 Consensus quality: 157449 bases at least Q20
 Insert size: 158066; sum-of-contigs
 Quality coverage: 7.99x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of
 41 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved
 1 16419 contig of 16419 bp in length
 16520 25963 contig of 9444 bp in length
 26064 37626 contig of 11563 bp in length
 37727 46930 contig of 9204 bp in length
 47031 52631 contig of 5601 bp in length
 52732 62656 contig of 9925 bp in length
 62757 70268 contig of 7512 bp in length
 70369 75300 contig of 4932 bp in length
 75401 77829 contig of 2429 bp in length
 77930 80472 contig of 2543 bp in length

*	103959	107722:	contlig	of	3764	bp	in	length
*	107723	107822:	gap	of	100	bp		
*	107823	112230:	contlig	of	4708	bp	in	length
*	112531	112630:	gap	of	100	bp		
*	112631	114591:	contlig	of	1961	bp	in	length
*	114592	114691:	gap	of	100	bp		
*	114692	116935:	contlig	of	224	bp	in	length
*	116936	117035:	gap	of	100	bp		
*	117036	119393:	contlig	of	228	bp	in	length
*	112294	119393:	gap	of	100	bp		
*	113994	124203:	contlig	of	3010	bp	in	length
*	124004	122503:	gap	of	100	bp		
*	125004	123781:	contlig	of	1278	bp	in	length
*	123782	123881:	gap	of	100	bp		
*	123882	126714:	contlig	of	283	bp	in	length
*	126715	126814:	gap	of	100	bp		
*	126815	129127:	contlig	of	233	bp	in	length
*	129128	129227:	gap	of	100	bp		
*	132128	132219:	contlig	of	2992	bp	in	length
*	132220	133319:	gap	of	100	bp		
*	133320	134600:	contlig	of	2441	bp	in	length
*	134601	134900:	gap	of	100	bp		
*	134901	137803:	contlig	of	2903	bp	in	length
*	137804	137903:	gap	of	100	bp		
*	137904	139584:	contlig	of	1661	bp	in	length
*	139585	139684:	gap	of	100	bp		
*	139685	143727:	contlig	of	4043	bp	in	length
*	143728	143827:	gap	of	100	bp		
*	143828	145778:	contlig	of	1951	bp	in	length
*	145779	145878:	gap	of	183	bp	in	length
*	145879	147711:	contlig	of	183	bp	in	length

*	147712	147811	gap of	100 bp
*	147812	148331	contig of	1120 bp
*	149332	149031	gap of	100 bp
*	149032	151135	contig of	2204 bp
*	151236	151335	gap of	100 bp
*	151336	153397	contig of	2552 bp
*	153998	153997	gap of	1714 bp
*	153998	155711	contig of	1714 bp

	*	15712	155811	gap of	100 bp	
	*	155812	157683	contig of	1872 bp	in length
	*	157684	157783	gap of	100 bp	in length
	*	157784	159523	contig of	1740 bp	in length
	*	159524	159623	gap of	100 bp	in length
	*	159624	160923	contig of	1300 bp	in length
	*	160924	161023	gap of	100 bp	in length
	*	161024	162066	contig of	1043 bp	in length
FEATURES						
source		1.	162066			
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/chromosome="11"				
		/map="11g"				
		/clone="RP11-121M22"				
		1.	16419			
misc_feature		/note="assembly_fragment"				
misc_feature		16520..25963				
misc_feature		/note="assembly_fragment"				
misc_feature		26064..37626				
misc_feature		/note="assembly_fragment"				
misc_feature		37727..46930				
misc_feature		/note="assembly_fragment"				
misc_feature		47031..52631				
misc_feature		/note="assembly_fragment"				
misc_feature		52732..62656				
misc_feature		/note="assembly_fragment"				
misc_feature		62757..70268				
misc_feature		/note="assembly_fragment"				
alignment_scores:						
	quality:	9.00		length:	9	
	ratio:	1.000		gaps:	0	
Percent Similarity:	100.000		Percent Identity:	100.000		

alignment block:
US-09-697-089-2 x AP002986/rev ..

Align seg 1/1 to reverse of: AP002986 from: 1 to: 162066

173 LysGlySerThreuleuglnArg 181
|||||
113714 AAGGAAATCCACTCTACTCAGCG 113688

seq_name: gb_htg:AC022488

seq_documentation_block:

LOCUS AC022488 162445 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-664F20, WORKING DRAFT
SEQUENCE, 13 unordered pieces.

ACCESSION AC022488
VERSION AC022488.2 GI:7105583

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human;

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 162445)
AUTHORS Waterston,R.H.

JOURNAL The sequence of Homo sapiens clone
TITLE Unpublished
2 (bases 1 to 162445)

REFERENCE Waterston,R.H.
AUTHORS

JOURNAL Direct Submission
TITLE Submitted (04-FEB-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT On Feb 25, 2000 this sequence version replaced gi:6899790.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0664F20
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big.Dye; 0% of reads
Chemistry: Dye-terminator Big.Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151216 bases at least Q40
Consensus quality: 155317 bases at least Q30
Consensus quality: 157348 bases at least Q20
Insert size: 18700; agarose-fp
Insert size: 161245; sum-of-contigs
Quality coverage: 4.17 in Q20 bases; agarose-fp
Quality coverage: 4.88 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 2226: contig of 2226 bp in length
* 2227 2326: gap of unknown length
* 2327 4009: contig of 1683 bp in length
* 4010 4109: gap of unknown length
* 4110 7025: contig of 2916 bp in length
* 7026 7125: gap of unknown length
* 7126 9060: contig of 1935 bp in length
* 9061 9161: gap of unknown length
* 9161 13139: contig of 3979 bp in length
* 13140 13239: gap of unknown length

13240 17695: contig of 4456 bp in length
* 17696 17795: gap of unknown length
* 17796 27844: contig of 10049 bp in length
* 27845 27944: gap of unknown length
* 27945 40720: contig of 12776 bp in length
* 40721 40821: gap of unknown length
* 40821 50885: contig of 10065 bp in length
* 50886 50985: gap of unknown length
* 50986 64989: contig of 14004 bp in length
* 64990 65089: gap of unknown length
* 65090 80120: contig of 15031 bp in length
* 80121 80220: gap of unknown length
* 80221 99954: contig of 19734 bp in length
* 99955 100054: gap of unknown length
* 100055 162445: contig of 62391 bp in length.
FEATURES
source
1. 162445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-664F20"
1. 2226
/note="assembly_name:Contig4"
2327. 4009
/note="assembly_name:Contig5"
4110. 7025
/note="assembly_name:Contig6"
7126. 9060
/note="assembly_name:Contig7"
9161. 13139
/note="assembly_name:Contig8"
13240. 17695
/note="assembly_name:Contig9"
17796. 27844
/note="assembly_name:Contig10
clone_end::77
vector_side:right"
27945. 40720
/note="assembly_name:Contig11"
40821. 50885
/note="assembly_name:Contig12"
50986. 64989
/note="assembly_name:Contig13"
65090. 80120
/note="assembly_name:Contig14
clone_end::SP6
vector_side:right"
80221. 99954
/note="assembly_name:Contig15"
100055. 162445
/note="assembly_name:Contig16"
BASE COUNT 39599 a 40899 c 40868 g 39847 t 1232 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AC022488 ..
Align seg 1/1 to: AC022488 from: 1 to: 162445
827 LeuValSerCysLeuSerAlaasn 835
|||||
49876 CTAGTCTTGTCTGCTGCTCTAAT 49902
seq_name: gb_pr:AC011489
seq_documentation_block:
LOCUS AC011489 164034 bp DNA PRI 04-AUG-2000

```

JOURNAL      unpublished
REFERENCE    2 (bases 1 to 165105)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
Project Information
Center Project Name: 527341
Center Clone name: RPC1-11_315A16
-----
Summary Statistics
Consensus quality: 145261 bases at least Q40
Consensus quality: 157158 bases at least Q30
Consensus quality: 159051 bases at least Q20
Estimated insert size: 168270; agarose-fp estimation
Estimated insert size: 163205; sum-of-contigs estimation
Quality coverage: 3.91 in Q20 bases; agarose-fp estimation
Quality coverage: 4.03 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1053: contig of 1053 bp in length
1054
1153: gap of unknown length
1154
2229: contig of 1076 bp in length
2230
2330
2330: gap of unknown length
3695: contig of 1366 bp in length
3696
3795: gap of unknown length
3796
5012: contig of 1217 bp in length
5013
5112: gap of unknown length
5113
7505: contig of 2393 bp in length
7506
7605: gap of unknown length
7606
10531: contig of 2926 bp in length
10532
10631: gap of unknown length
10632
12441: contig of 1810 bp in length
12442
12541: gap of unknown length
12542
15856: contig of 3315 bp in length
15857
15956: gap of unknown length
15957
19342: contig of 3386 bp in length
19343
19442: gap of unknown length
19443
23101: contig of 3659 bp in length
23102
23201: gap of unknown length
23202
26538: contig of 3337 bp in length
26539
26638: gap of unknown length
26639
30342: contig of 3704 bp in length
30343
30442: gap of unknown length
30443
35600: contig of 5158 bp in length
35601
35700: gap of unknown length
35701
44202: contig of 8502 bp in length
44203
44302: gap of unknown length
44303
55754: contig of 11452 bp in length
55755
55854: gap of unknown length
55855
71827: contig of 15973 bp in length
71828
71927: gap of unknown length
71928
90624: contig of 18697 bp in length
90625
90724: gap of unknown length
90725
111439: contig of 20715 bp in length
111440
111539: gap of unknown length
111540
132012: contig of 20473 bp in length
132013
132112: gap of unknown length
132113
165105: contig of 32993 bp in length.
Location/Qualifiers
1..165105 "Homo sapiens"
/organism="Homo sapiens"
source

```

/db_xref="taxon:9606"
 /chromosome="5"
 /clone_lib="RP11-315A16"
 /clone="31184 c 30734 g 51092 c 1915 others
 BASE COUNT 50180 a 31184 c 30734 g 51092 c 1915 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AC091932/rev ..

Align seg 1/1 to reverse of: AC091932 from: 1 to: 165105

809 AspyrilleValleySerleusSer 817
 ||||||||||||||||||
 75732 GACTACATGTGCAAGCTGTCAACT 75706

seq_name: gb_htg:AC025884

seq_documentation_block:

LOCUS AC025884 165341 bp DNA HTG 21-JUL-2001
 DEFINITION Homo sapiens chromosome 15 clone RP11-603B24 map 15, WORKING DRAFT
 SEQUENCE, 6 unordered pieces.

AC025884
 AC025884.4 GI:14971329
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AITHONS
 TITLE
 JOURNAL
 REFERENCE
 AITHONS
 1 (bases 1 to 165341)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 15, clone RP11-603B24
 Unpublished
 2 (bases 1 to 165341)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
 Boguski, M., Bouckge, K., Brown, A., Burdett, G.,
 Campiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearlano, R., Dewar, K., Diaz, J. S.,
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, D. C., Iliev, I., Johnson, R., Jones, C., Kahn, L., Karatas, A.,
 Klein, J., Lakoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheters, R.,
 Melchior, J., Menus, L., Mihov, T., Miranda, C., Mleca, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zannou, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 20, 2001 this sequence version replaced gl:13959264.
 All repeats were identified using RepeatMasker:
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: W1BR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L8207
 Center clone name: 603.B.24

----- Summary Statistics
 Sequencing vector: M13; M77815; 35% of reads
 Sequencing vector: Plasmid; n/a; 65% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; Version 0.960731
 Consensus quality: 163980 bases at least Q40
 Consensus quality: 164684 bases at least Q30
 Consensus quality: 164684 bases at least Q20
 Insert size: 167000; agarose-fp
 Insert size: 164841; sum-of-contigs
 Quality coverage: 12.9 in Q20 bases; agarose-fp
 Quality coverage: 13.1 in Q20.

* NOTE: This is a working draft sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 19796: contig of 19796 bp in length
 * 19797 19896: gap of 100 bp
 * 19897 23952: contig of 4056 bp in length
 * 23953 24052: gap of 100 bp
 * 24053 40348: contig of 16296 bp in length
 * 40349 40448: gap of 100 bp
 * 40449 79842: contig of 38394 bp in length
 * 79843 79942: gap of 100 bp
 * 79943 132333: contig of 52391 bp in length
 * 132334 132433: gap of 100 bp
 * 132434 165341: contig of 32908 bp in length.

FEATURES

SOURCE

1. 165341

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"

/clone="RP11-603B24"

/clone_lib="RPCT-11 Human Male BAC"

1. 19796

/note="assembly-fragment"

clone_end:SP6

vector_side:left"

19897. 23952

/note="assembly-fragment"

24053. 40348

/note="assembly-fragment"

40449. 79842

/note="assembly-fragment"

79943. 132333

/note="assembly-fragment"

132434. 165341

/note="assembly-fragment"

clone_end:17

vector_side:right"

BASE COUNT 47100 a 36508 c 36616 g 44616 t 501 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AC025884/rev ..

Align seg 1/1 to reverse of: AC025884 from: 1 to: 165341

895 SerleusSerleusSerleus 903

104369 TCACCTCTACCTCTCAACATCTG 104343

seq_name: gb_htg:AC060768

seq_documentation_block:
LOCUS AC060768 165524 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-57613 map 2, WORKING DRAFT
SEQUENCE 39 unordered pieces.
AC060768
VERSION AC060768.2 GI:8097399
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 2, clone RP11-57613
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165524)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
Galand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McNeeters, R.,
Melgrim, J., Menus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Plante, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tittell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT Submitted (20-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 26, 2000 this sequence version replaced 91:7622388.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information

Center project name: L6360
Center clone name: 5761_3

Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 144591 bases at least Q40
Consensus quality: 154311 bases at least Q30
Consensus quality: 158850 bases at least Q20
Insert size: 170000; agarose-fp
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 1113: contig of 1113 bp in length
1114 1213: gap of 100 bp
1214 2580: contig of 1367 bp in length
2581 2680: gap of 100 bp
2681 3936: contig of 1256 bp in length
3937 4036: gap of 100 bp
4037 5112: contig of 1076 bp in length
5113 5212: gap of 100 bp
5213 6383: contig of 1171 bp in length
6384 6483: gap of 100 bp
6484 8013: contig of 1530 bp in length
8014 8113: gap of 100 bp
8114 10576: contig of 2463 bp in length
10577 10676: gap of 100 bp
10677 12173: contig of 1497 bp in length
12174 12273: gap of 100 bp
12274 13512: contig of 1239 bp in length
13513 13612: gap of 100 bp
13613 16041: contig of 2429 bp in length
16042 16141: gap of 100 bp
16142 18208: contig of 2067 bp in length
18209 18308: gap of 100 bp
18309 19930: contig of 1622 bp in length
19931 20030: gap of 100 bp
20031 22516: contig of 2486 bp in length
22517 22616: gap of 100 bp
22617 24800: contig of 2184 bp in length
24801 24900: gap of 100 bp
24901 27062: contig of 2162 bp in length
27063 27162: gap of 100 bp
27163 28527: contig of 1365 bp in length
28528 28627: gap of 100 bp
28628 31113: contig of 2486 bp in length
31114 31213: gap of 100 bp
31214 33582: contig of 2369 bp in length
33583 33682: gap of 100 bp
33683 37785: contig of 4103 bp in length
37786 37885: gap of 100 bp
37886 40875: contig of 2990 bp in length
40876 40975: gap of 100 bp
40976 43294: contig of 2319 bp in length
43295 43394: gap of 100 bp
43395 45738: contig of 2344 bp in length
45739 45838: gap of 100 bp
45839 47771: contig of 1933 bp in length
47772 47871: gap of 100 bp
47872 49090: contig of 1219 bp in length
49091 49190: gap of 100 bp
49191 52213: contig of 3033 bp in length
52214 52313: gap of 100 bp
52314 54678: contig of 2365 bp in length
54679 54778: gap of 100 bp
54779 58759: contig of 3961 bp in length
58760 58859: gap of 100 bp
58860 63379: contig of 4530 bp in length
63380 63479: gap of 100 bp
63480 68488: contig of 5009 bp in length
68489 68588: gap of 100 bp
68589 74028: contig of 5440 bp in length
74029 74128: gap of 100 bp
74129 80264: contig of 6136 bp in length
80265 80364: gap of 100 bp
80365 86363: contig of 5999 bp in length
86364 86463: gap of 100 bp
86464 93557: contig of 7094 bp in length
93558 93657: gap of 100 bp
93658 102460: contig of 8803 bp in length
102461 102560: gap of 100 bp
102561 110468: contig of 7908 bp in length


```

BASE COUNT      51633 a 31975 c 32631 g 49531 t 269 others
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000     Gaps:      0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x ACO21915/rev ..

Align seg 1/1 to reverse of: ACO21915 from: 1 to: 166039

483 11eThSeThrTySeSerLeuLeu 491
|||||
101574 ATTACAGCAGCTTATTCAGCCTCTTT 101548

seq_name: gb_htg:ACO80132

seq_documentation_block:
LOCUS      ACO80132      167839 bp      DNA      HTG      24-FEB-2001
DEFINITION Homo sapiens chromosome 4 clone RP11-166B20, WORKING DRAFT
SEQUENCE, 29 unordered pieces.
ACCESSION  ACO80132
VERSION     ACO80132.4 GI:13122785
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUZZTOP.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 167839)
Waterston,R.H.
The sequence of Homo sapiens clone
unpublished
2 (bases 1 to 167839)
Waterston,R.H.
Direct Submission
Submitted (25-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis
MO 63108 USA
On Feb 24, 2001 this sequence version replaced gi:11094870.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0166B20
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: Plasmid, 0%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 152050 bases at least Q40
Consensus quality: 157361 bases at least Q30
Consensus quality: 159834 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 165039; sum-of-contigs
Quality coverage: 3.67 in Q20 bases; agarose-fp
Quality coverage: 3.78 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1324: contig of 1324 bp in length

```

*	1325	1424: gap of unknown length
*	1425	2848: contig of 1424 bp in length
*	2849	2948: gap of unknown length
*	2949	4316: contig of 1368 bp in length
*	4317	4416: gap of unknown length
*	4417	5995: contig of 1579 bp in length
*	5996	6095: gap of unknown length
*	6096	7813: contig of 1718 bp in length
*	7814	7913: gap of unknown length
*	10009	10008: contig of 2095 bp in length
*	10109	12161: contig of 2053 bp in length
*	12162	12261: gap of unknown length
*	12262	14309: contig of 2048 bp in length
*	14310	14409: gap of unknown length
*	14410	18142: contig of 3733 bp in length
*	18143	18242: gap of unknown length
*	18243	21181: contig of 2939 bp in length
*	21182	21281: gap of unknown length
*	21282	23710: contig of 2429 bp in length
*	23711	23810: gap of unknown length
*	23811	27376: contig of 3466 bp in length
*	27377	27376: gap of unknown length
*	27377	28657: contig of 1281 bp in length
*	28658	28757: gap of unknown length
*	28758	34079: contig of 5332 bp in length
*	34080	34179: gap of unknown length
*	34180	38342: contig of 4153 bp in length
*	38343	38442: gap of unknown length
*	38443	43780: contig of 5338 bp in length
*	43781	43880: gap of unknown length
*	43881	46597: contig of 5817 bp in length
*	46598	49797: gap of unknown length
*	49798	55866: contig of 6069 bp in length
*	55867	55966: gap of unknown length
*	55967	61985: contig of 6019 bp in length
*	61986	62085: gap of unknown length
*	62086	68007: contig of 5922 bp in length
*	68008	68107: gap of unknown length
*	73474	73474: contig of 5367 bp in length
*	73475	73574: gap of unknown length
*	73575	82789: contig of 9215 bp in length
*	82790	82889: gap of unknown length
*	82890	91137: contig of 8848 bp in length
*	91738	91837: gap of unknown length
*	91838	99614: contig of 7777 bp in length
*	99615	99714: gap of unknown length
*	99715	104002: contig of 4288 bp in length
*	104003	104102: gap of unknown length
*	104103	118611: contig of 14509 bp in length
*	118612	118711: gap of unknown length
*	118712	132662: contig of 13951 bp in length
*	132663	132762: gap of unknown length
*	132763	148768: contig of 16006 bp in length
*	148769	148868: gap of unknown length
*	148869	167839: contig of 18971 bp in length
FEATURES		
source		
	1. 167839	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="4"	
	/clone="RP11-166B20"	
misc_feature	1. 1324	
	/note="assembly_name:Contig15"	
misc_feature	1425. 2848	
	/note="assembly_name:Contig16"	
misc_feature	2949. 4316	
	/note="assembly_name:Contig17"	
misc_feature	4417. 5995	
	/note="assembly_name:Contig18"	
misc_feature	6096. 7813	
	/note="assembly_name:Contig19"	
misc_feature	7914. 10008	

```

misc_feature /note="assembly_name:Contig20"
10109..12161
/note="assembly_name:Contig21"
misc_feature /note="assembly_name:Contig21"
12262..14309
/note="assembly_name:Contig22"
misc_feature /note="assembly_name:Contig22"
14410..18142
/note="assembly_name:Contig23"
misc_feature /note="assembly_name:Contig23"
18243..21181
/note="assembly_name:Contig24"
misc_feature /note="assembly_name:Contig24"
21282..23710
/note="assembly_name:Contig25"
misc_feature /note="assembly_name:Contig25"
23811..27276
/note="assembly_name:Contig26"
misc_feature /note="assembly_name:Contig26"
27377..28657
/note="assembly_name:Contig27"
misc_feature /note="assembly_name:Contig27"
28758..34079
/note="assembly_name:Contig28"
misc_feature /note="assembly_name:Contig28"
34180..38342
/note="assembly_name:Contig29"
misc_feature /note="assembly_name:Contig29"
38443..43780
/note="assembly_name:Contig30"
misc_feature /note="assembly_name:Contig30"
43881..49697
/note="assembly_name:Contig31"
misc_feature /note="assembly_name:Contig31"
49798..55866
/note="assembly_name:Contig32"
misc_feature /note="assembly_name:Contig32"
53967..61985
/note="assembly_name:Contig33"
misc_feature /note="assembly_name:Contig33"
62086..68007
/note="assembly_name:Contig34"
misc_feature /note="assembly_name:Contig34"
68108..73474
/note="assembly_name:Contig35"
misc_feature /note="assembly_name:Contig35"
73575..82789
/note="assembly_name:Contig36"
misc_feature /note="assembly_name:Contig36"
82890..91737
/note="assembly_name:Contig37"
misc_feature /note="assembly_name:Contig37"
91838..99614
/note="assembly_name:Contig38"
misc_feature /note="assembly_name:Contig38"
99715..104002
/note="assembly_name:Contig39"
misc_feature /note="assembly_name:Contig39"
104103..118611
/note="assembly_name:Contig40"
misc_feature /note="assembly_name:Contig40"
118712..132662
/note="assembly_name:Contig41"
misc_feature /note="assembly_name:Contig41"
132763..148768
/note="assembly_name:Contig42"
misc_feature /note="assembly_name:Contig42"
148869..167839
/note="assembly_name:Contig43"
BASE COUNT 53172 a 29418 c 29378 g 53062 t 2809 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AC080132/rev ..
Align seg 1/1 to reverse of: AC080132 from: 1 to: 167839
126 llellepheasnlulysertlrphe 134
|||||
133050 ATTATCTTCACCTGACGTCACGTTT 133024
seq_name: gb_pr:AC024367
seq_documentation_block:
LOCUS AC024367 170022 bp DNA PRI 30-MAY-2001
DEFINITION Homo sapiens chromosome 8, clone RP11-75G10, complete sequence.
ACCESSION AC024367
VERSION AC024367.6 GI:14249088
KEYWORDS HTG.

```

```

SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone RP11-75G10
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 170022)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bedalov, B., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Burkett, G., Campione, A., Castle, A.,
Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Donato, M., Doyle, M.,
Fenstermaker, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M.,
McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Miranda, C., Mianga, V., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
Travers, W., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.
JOURNAL Direct Submission
AUTHORS Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 170022)
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
Camata, J., Campione, A., Chang, J., Choquel, Y., Colangelo, M.,
Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K.,
Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,
Iliev, I., Johnson, R., Jones, C., Karatas, A., Larocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Margulis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L.,
Mihova, T., Mianga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,
Theodore, J., Travers, W., Trays, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
JOURNAL Direct Submission
AUTHORS Submitted (30-MAY-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 30, 2001 this sequence version replaced gi:1414187.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6126
Center clone name: 75_G_10
----- Location/Qualifiers
FEATURES
Source 1..170022

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-75G10"
/clone_lib="RPC1-11 Human Male BAC"
6. .135
repeat_region /rpt_family="MER21B"
267. .436
repeat_region /rpt_family="(TG)n"
494. .616
repeat_region /rpt_family="(TG)n"
2560. .2638
repeat_region /rpt_family="MER116"
3995. .4016
repeat_region /rpt_family="AT-rich"
4164. .4473
repeat_region /rpt_family="AluY"
complement(4487. .4816)
/rpt_family="L1R16B"
complement(5272. .5476)
/rpt_family="MIR"
5491. .5841
repeat_region /rpt_family="MLT1A2"
6998. .7392
repeat_region /rpt_family="MSTD"
7738. .7819
repeat_region /rpt_family="AluSp/g"
complement(7820. .7871)
/rpt_family="L1M4"
complement(7872. .7886)
/rpt_family="AluSg"
7887. .7910
repeat_region /rpt_family="(GTTTG)n"
complement(7911. .8155)
/rpt_family="AluSg"
complement(8275. .8515)
/rpt_family="L3"
9260. .9329
repeat_region /rpt_family="L2"
10622. .10792
repeat_region /rpt_family="MER5B"
10799. .11098
repeat_region /rpt_family="AluSx"
complement(12289. .12407)
/rpt_family="L2"
12449. .12520
repeat_region /rpt_family="MIR"
12665. .12893
repeat_region /rpt_family="L2"
13156. .13259
repeat_region /rpt_family="L2"
13369. .13701
repeat_region /rpt_family="MLT1H"
complement(13966. .14263)
/rpt_family="Aluub"
complement(14933. .15101)
/rpt_family="MIR"
16078. .16200
repeat_region /rpt_family="L2"
16457. .16485
repeat_region /rpt_family="AT-rich"
16956. .17240
repeat_region /rpt_family="Aluub"
17250. .17272
repeat_region /rpt_family="(CAAA)n"
17307. .17486
repeat_region /rpt_family="MIR"
17681. .17936
repeat_region /rpt_family="Charlie2"
17941. .18800
repeat_region /rpt_family="Charlie2"
19617. .20038

```

```

repeat_region /rpt_family="L2"
complement(20262. .20444)
/rpt_family="MER53"
21120. .21159
repeat_region /rpt_family="AT-rich"
21412. .21719
repeat_region /rpt_family="L1MC5"
21822. .21868
repeat_region /rpt_family="(CA)n"
complement(21914. .22136)
/rpt_family="MIR"
22422. .22452
repeat_region /rpt_family="(TTTGTG)n"
complement(23088. .23331)
/rpt_family="MIR"
complement(23507. .23711)
/rpt_family="MIR"
23733. .23956
repeat_region /rpt_family="MIR"
24109. .24195
repeat_region /rpt_family="A-rich"
24252. .24470
/rpt_family="MIR"
25283. .25569
repeat_region /rpt_family="AluSx"
26614. .26780
repeat_region /rpt_family="FRAM"
complement(27083. .27277)
/rpt_family="L1McA"
27343. .27676
repeat_region /rpt_family="MLT1P1"
27677. .27979
repeat_region /rpt_family="Aluub"
27980. .28098
repeat_region /rpt_family="MLT1P1"
29239. .29301
/rpt_family="MLT1P"
complement(31283. .31557)
/rpt_family="AluSc"
complement(31577. .31641)
/rpt_family="MER5B"
31642. .31851
repeat_region /rpt_family="L1P81"
complement(31859. .31996)
/rpt_family="MER5B"
complement(32141. .32673)
/rpt_family="MLT1P1"
32674. .32703
repeat_region /rpt_family="AT-rich"

```

```

alignment_scores:
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

us-09-697-089-2 x AC024367/rev ..

Align seg 1/1 to reverse of: AC024367 from: 1 to: 170022

```

414 SerservAlasngluspVallleu 422
|||||
57030 TCCTCAGTAATGAGATGCTCTTTG 57004

```

THIS PAGE BLANK (USPTO)